

HUMBOLDT-UNIVERSITÄT ZU BERLIN

INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 443

DATA STEWARDSHIP ALS BOUNDARY-WORK

VON
LAURA ROTHFRITZ

DATA STEWARDSHIP ALS BOUNDARY-WORK

VON
LAURA ROTHFRITZ

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Vivien Petras
Humboldt-Universität zu Berlin

Heft 443

Rothfritz, Laura

Data Stewardship als Boundary-Work / von Laura Rothfritz. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2019. – 125 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 443)

ISSN 14 38-76 62

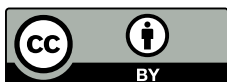
Abstract

Data Stewardship wird besonders im Bezug auf die FAIR Data Principles und die Einführung der European Open Science Cloud als Begriff weitläufig verwendet, der im weitesten Sinne alle anfallenden Aktivitäten im Forschungsdatenmanagement beschreibt. Forschungsdatenmanagement kann als soziotechnisches System angesehen werden, in dem unterschiedliche Akteure in Beziehung treten und zusammenwirken müssen. Die Akteure entstammen dabei unterschiedlichen Sozialen Welten, die an ihren Schwellen aufeinandertreffen und Grenzbereiche (*Boundaries*) bilden. Unterschiedliche Methoden der Grenzarbeit (*Boundary-Work*) können angewendet werden, um Kollaborationen zwischen Akteursgruppen zu ermöglichen und zu optimieren.

In dieser Arbeit wurden Theorien zu Kollaborationen in soziotechnischen Systemen und des Agierens an Boundaries aus der Technik- und Wissenschaftssoziologie auf das Anwendungsfeld des Data Stewardships übertragen. Hierfür wurden begriffliche Entwicklungen und Ausprägungen des Konzeptes Data Stewardship nachvollzogen und bestehende theoretische Ansätze zu Methoden und Aufgaben von Boundary-Work auf die Domäne Data Stewardship angewendet. Diese Arbeit ermöglicht eine neue Perspektive auf Data Stewardship Aktivitäten und die Rolle von Data Stewards. Sie zeigt, dass Data Stewardship als ein Zusammenspiel unterschiedlicher Akteure eingebettet in soziale Geflechte angesehen werden kann. Data Stewards müssen dabei über infrastrukturelle oder forschungsgetriebene Zielsetzungen hinaus vermittelnd auf einer sozialen Ebene agieren und dafür entsprechende Kompetenzen mitbringen.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im Studiengang Informationswissenschaften M.A. an der Fachhochschule Potsdam.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.



Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 4.0 International zugänglich. Eine Kopie dieser Lizenz finden Sie unter <http://creativecommons.org/licenses/by/4.0/>

Inhaltsverzeichnis

1	Einleitung	9
1.1	Motivation	9
1.2	Forschungsfragen	12
2	Forschungsstand	14
2.1	Neuere Entwicklungen im Forschungsdatenmanagement	14
2.2	Theoretische Grundlagen	19
3	Methode	23
4	Data Stewardship	25
4.1	Begriff und Bedeutungsentwicklung	25
4.2	Begriffsspektrum	36
4.3	Verwendung des Begriffs in dieser Arbeit	41
5	Boundary Work	42
5.1	Arten von Boundary Work	42
5.2	Arten von Boundaries	44
5.3	Verwendung von Boundary Objects	45
6	Boundaries im Data Stewardship	50
6.1	Abgrenzung der betrachteten Akteure	50
6.2	Data Stewardship Tätigkeiten an Boundaries	53
7	Data Stewardship als Boundary Work	71
7.1	Akteure und Boundaries	71
7.2	Data Stewards als Boundary Spanner	74
7.3	Mögliche Boundary Objects	76
8	Fazit und Ausblick	79
	Literatur	85

Anhang	95
---------------	-----------

Anhang A Tätigkeiten und Akteure	96
---	-----------

A.1 Plan and design	96
A.2 Capture and process	99
A.3 Integrate and analyse	102
A.4 Appraise and preserve	106
A.5 Publish and release	109
A.6 Expose and discover	112
A.7 Govern and assess	115
A.8 Scope and resource	118
A.9 Advise and enable	122

Abkürzungsverzeichnis

BMBF	Bundesministerium für Bildung und Forschung
BO	Boundary Object
DFG	Deutsche Forschungsgemeinschaft
DMP	Datenmanagementplan
DTL	Dutch Techcenter for Lifesciences
NFDI	Nationale Forschungsdateninfrastruktur
EOSC	European Open Science Cloud
ESDSF	Enterprise Scientific Data Stewardship Framework
HRK	Hochschulrektorenkonferenz
LCRDM	Landelijk Coördinatiepunt Research Data Management (nationale Forschungsdaten-Koordinationsstelle, Niederlande)
LTER	Long Term Ecological Research
LZA	(Digitale) Langzeitarchivierung
NPM	New public management
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
OA	Open Access
PID	Persistent Identifier
RfII	Rat für Informationsinfrastrukturen
RDA	Research Data Alliance
STS	Science and Technology Studies
WR	Wissenschaftsrat

Tabellenverzeichnis

1	Begriffsentwicklung Stewardship	35
2	Unterschiedliche Arten von Knowledge Boundaries	46
3	Tätigkeiten „Plan and design“	56
4	Tätigkeiten „Capture and process“	57
5	Tätigkeiten „Integrate and analyse“	59
6	Tätigkeiten „Appraise and preserve“	60
7	Tätigkeiten „Publish and release“	62
8	Tätigkeiten „Expose and discover“	64
9	Tätigkeiten „Govern and assess“	65
10	Tätigkeiten „Scope and ressource“	67
11	Tätigkeiten „Advise and enable“	69
12	Akteure und Anzahl der Tätigkeiten mit Beteiligung	71
13	Kategorien von Boundaries und identifizierte Anzahlen	73

Abbildungsverzeichnis

1	Scope des Data Stewardships für Umweltdaten	31
2	Darstellung der Stewardship Dimensionen des DTL	34
3	Aufgabenbereiche im Data Stewardship	38
4	Knowledge Boundaries	45
5	Planning data management and sharing (DMP)	96
6	Open data model and database design.	97
7	Metadata, persistent ID specification.	97
8	Open source software / service requirements	98
9	Repository and data management plattform appraisal.	98
10	Workflow setup and documentation	99
11	Database management	99
12	Software prototyping	100
13	Data collection and reuse of open data	100
14	File naming and organisation	101
15	Data provenance and software versioning	101
16	Math and statistical knowledge application	102
17	Critical thinking and theory building	102
18	Creative problem solving, flexibility	103
19	Open source software / service development	103
20	Data transformation and integration	104
21	Data mining, querying, interpretation	104
22	Predictive modelling and analytics	105
23	Machine learning methods	105
24	Data preparation, documentation for reproducibility	106
25	Data quality assurance using open standards	106

26	Ethical, legal and data policy compliance	107
27	Data transfer and long-term storage	107
28	Format and media migration	108
29	Software review and preservation	108
30	Documentation for public use, lay understanding	109
31	Access control and management	109
32	Ethical application of patents, licenses	110
33	Open access publishing and self-archiving	110
34	Sharing via open repositories / platforms	111
35	Engaging in open innovation beyond academia	111
36	Vocabulary / ontology application	112
37	Metadata and PID exposure	112
38	Visualization and presentation of results	113
39	Evaluation of repository and publishing platforms	113
40	Repository / database search	114
41	Citation of research outputs	114
42	Research strategy / open research vision	115
43	Advocacy of FAIR principles and Open Access policy	115
44	Research integrity, attribution, impact awareness	116
45	Information security and risk management	116
46	Data governance, handling third-party data	117
47	Storage security management	117
48	Contributing to quality assessment or peer review	118
49	Secure funding for open science / support	118
50	Perseverance delivering diverse open research	119
51	Service level management	119
52	Change management	120
53	Workflow set-up and provenance information management	120
54	Cloud environment and storage management	121
55	Authentication and authorisation (AAI) management	121
56	Costing of data management and preservation	122
57	Building open inter-disciplinary collaborations	122
58	Engaging with research users and stakeholders	123
59	Developing a profile of open research	123
60	Training in open methods, services	124
61	Contributing to education, professional development	124
62	Contributing to open RES, networks, standards bodies	125
63	Supervision and mentoring	125

1 Einleitung

1.1 Motivation

Das Management von Forschungsdaten im Zuge einer transparenten, nachvollziehbaren und replizierbaren Forschung hat in den letzten Jahren zunehmend an Bedeutung gewonnen. Mit der Entwicklung von nationalen Infrastrukturen, wie der Nationalen Forschungsdateninfrastruktur (NFDI) in Deutschland und international übergreifenden Aktivitäten, wie der European Open Science Cloud (EOSC), bleiben Fragen zu Methoden, Technologien und Rahmenbedingungen im Umgang mit Forschungsdaten weiterhin aktuell. Zu Aufgaben, Verantwortungsverteilung und technischen Desideraten für das Management von Forschungsdaten erschienen in den letzten Jahren eine Reihe von Studien, die beispielsweise die Rolle der Bibliotheken (z.B. Tenopir et al., 2014), Haltungen und Praktiken von einzelnen Fachdisziplinen (z.B. Abele-Brehm et al., 2019) und ganzen Institutionen (z.B. Kindling, Schirmbacher und Simukovic, 2013) oder Motivationen und Hemmnisse beschreiben (z.B. Fecher et al., 2017; Tammaro et al., 2017).

Unter das Forschungsdatenmanagement fallen eine Anzahl von Datenmanagementaktivitäten sowie die ihnen übergeordneten Begrifflichkeiten der Kuratierung (*Curation*) und der Erhaltung (*Preservation*) von Forschungsdaten, die wie folgt in dieser Arbeit verstanden werden:

- **Curation** beschreibt alle Maßnahmen, die während der „aktiven“ Phase der Datennutzung und -bereitstellung statt finden. Dabei handelt es sich insbesondere um die Übernahme von Daten in ein Repositorium, um die Beschreibung mit Metadaten zur Nachnutzung und Auffindbarkeit und um die Veröffentlichung (Publikation) von Datensätzen.
- **Preservation** bezeichnet die Langzeitarchivierung von Daten. Darunter wird die Übernahme in ein geeignetes Repositorium, die Beschreibung mit Metadaten zur Archivierung, sowie die Speicherung und ggf. die Transformierung und/oder Migration von Daten, verstanden.

Curation und Preservation werden in dieser Arbeit als sich gegenseitig bedingende jedoch voneinander distinkte Aktivitäten im Forschungsdatenmanagement verwendet. Forschungsdatenmanagement wird als „aktives“ Forschungsdatenmanagement und somit dynamischer Prozess, welches den Lebenszyklus von Daten von der Planung, ihrer Erhebung bis hin zu ihrer Publikation und Langzeitarchivierung begleitet verstanden (Neuroth et al., 2018). In der Literatur zum Thema Forschungsdatenmanagement lag der Fokus der Untersuchungen in den letzten Jahren besonders auf Technologien und Praktiken einzelner beteiligter Akteursgruppen, wie Forschende (z.B. Pels et al., 2018) oder Anforderungen für die Infrastruktur (z.B. Vogl, Rudolph und Thoring, 2019). Zudem wurden für die operative Durchführung eine Reihe von Konzepten zu Trainings- und Beratungsangeboten entwickelt, die sich im weitesten Sinne auf die traditionelle Rolle der Bibliotheken bei der Vermittlung von Informationskompetenz beziehen und die zum Ziel haben, Forschenden neue Formen von Kompetenzen (z.B. im Sinne einer Data Literacy) zu vermitteln (z.B. Dolzycka et al., 2019).

Unterschiedliche Akteure haben Anteil am Forschungsdatenmanagement. Häufig genannt werden Forschende selbst, Infrastruktureinrichtungen (wie Bibliotheken und Rechenzentren) und Forschungsförderer oder andere Akteure aus der Wissenschaftsadministration. Die Aufgaben, die von Seiten der Infrastrukturanbieter anfallen, werden in der Literatur als Forschungsdatenmanagement (*Research Data Management*)¹ (z.B. Büttner, Hobohm und Müller, 2011), Datenkuratierung (*Digital/Data Curation*) (z.B. Higgins, 2011), digitale Langzeitarchivierung (*Digital Preservation*) (z.B. Moore, 2008) oder im Englischen zusammenfassend häufig als *Long-term Stewardship* (Friedlander und Adler, 2006) bezeichnet. Besonders in der Entwicklung der EOSC und einem damit verbundenen Desiderat nach Daten, die so vorliegen, dass sie auffindbar, zugänglich, interoperabel und langfristig nachnutzbar sind, wurde in den letzten Jahren der Begriff *Data Stewardship* zunehmend verwendet. Was genau diese Tätigkeiten subsumieren und inwiefern sie sich von bereits etablierten Tätigkeiten der Infrastrukturanbieter unterscheiden, wird derzeit erforscht. Datenspezialist*innen sind für den Aufbau von nationalen und internationalen vernetzten Dateninfrastrukturen unerlässlich. Welche Kompetenzen Data Stewards mitbringen müssen, und welche Formen des Zusammenwirkens unterschiedlicher Akteure sie unterstützen können, ist noch nicht abschließend geklärt.

Data Stewardship ist ein englischer Begriff, für den es im Deutschen bislang keine Übersetzung gibt. Der Begriff kann auf unterschiedliche Arten verstanden werden. Sehr häufig wird der Begriff nicht präzise definiert, wenn in der Literatur von Stewardship die Rede ist. Jedoch häuft sich seine Verwendung auch außerhalb des englischsprachigen Raumes, z.B. durch den Aufbau von Data Stewardship Programmen und der Forderung nach einer vermehrten Ausbildung von Data Stewards. Beispielsweise müssten für ein „gutes“ Datenmanagement innerhalb aller Projekte auf europäischer Ebene laut einer Schätzung bis zu 500.000 Data Stewards ausgebildet werden (Versweyveld, 2016). Die niederländische Technische Universität Delft hat in Europa als Vorreiter in der Hochschullandschaft ein Data Stewardship Programm gestartet und beschäftigt Personen, die analog die Berufsbezeichnung *Data Steward* tragen. Diese Data Stewards werden von der Bibliothek der TU Delft zentral koordiniert und sind für jeweils ein Forschungsgebiet an der Hochschule zuständig.² Zwar gibt es hierzu eine Reihe von Erfahrungsberichten und Berichte zum Aufbau des Programms³, jedoch fehlen bislang längerfristig durchgeführte Studien zu genauen Praktiken und Wirken der Data Stewards. Das Projekt kann daher als ein Pilotprojekt betrachtet werden, an das die weitere Forschung anknüpfen kann.

Ein Ziel von Forschungsdatenmanagement-Aktivitäten ist, Daten auffindbar, zugänglich, interoperabel und nachnutzbar zu veröffentlichen. Diese in den *FAIR Data Principles* (Wilkinson et al., 2016) formulierten Qualitäten von (offenen) Forschungsdaten können jedoch nur durch eine Kollaboration unterschiedlicher Akteure erreicht werden. Gleichzeitig agieren alle Akteure auch mit technischen Systemen, die Daten erheben, verarbeiten und speichern

¹Im Folgenden werden englischsprachige Begriffe nach ihrer Einführung dem deutschen Sprachgebrauch folgend groß geschrieben, sofern es sich um Substantive oder Substantivierungen handelt. Begriffe werden bei ihrer Einführung *kursiv* als Fremdbegriff kenntlich gemacht und im weiteren Textverlauf nicht mehr gesondert markiert.

²Siehe: <https://www.tudelft.nl/en/library/current-topics/research-data-management/r/data-stewardship/>. Abgerufen am 21. Oktober 2019.

³Siehe: <https://openworking.wordpress.com/data-stewardship/>. Abgerufen am 21. Oktober 2019.

oder sie verwenden Daten als technisch basierte Abbildung der Realität. Hierbei kann es zu Wissensdivergenzen zwischen den Akteuren kommen, die sich als Grenze zwischen den unterschiedlichen Akteursgruppen zeigen können, und die hinderlich für eine optimale Zusammenarbeit sind. In diesem Sinne kann Forschungsdatenmanagement auch als ein soziotechnisches System (nach Trist und Bamforth, 1951) angesehen werden, in dem unterschiedliche Bereiche des Wissenschaftsbetriebs aufeinander treffen und in dem die Vertreter dieser Bereiche kollaborativ und technikgestützt agieren müssen.

Studien zu sozialen Aspekten bei der Zusammenarbeit im Umgang mit Forschungsdaten und welche Aufgaben auf Data Stewards zukommen, um diese Zusammenarbeit optimal zu gestalten, fehlen bislang weitgehend. Interaktionsprozesse in soziotechnischen Systemen werden unter dem Aspekt der Zusammenarbeit unterschiedlicher Akteursgruppen, die in ihre jeweiligen Sozialen Welten eingebunden sind betrachtet (Star und Griesemer, 1989). Soziale Welten treffen sich dabei an Schwellen oder Grenzen, die englisch als *Boundaries* bezeichnet werden.⁴ Ziel von Kollaborationen ist häufig das Erschaffen eines gemeinsamen Produkts oder eines neuen, optimierten Arbeitsmodus (z.B. in Bechky, 2003; Carlile, 2002; Kellogg, Orlikowski und Yates, 2006). Wissendivergenzen zwischen den unterschiedlichen Akteuren, in Form von *Knowledge Boundaries* können diesen Prozess behindern (Carlile, 2002). Für die Überwindung solcher *Boundaries* eignen sich daher koordinierende Schwellentätigkeiten (*Boundary-Work*), die von menschlichen Akteuren ausgeführt werden können. Hilfsmittel dafür kann die Verwendung von bestimmten Objekten oder Artefakten (*Boundary Objects*) sein.

Im Juli 2019 erschienen die Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft des Rats für Informationsinfrastrukturen (RfII). In diesen wird beschrieben, wie sich neuartige Kompetenzen und Berufsfelder in der Zusammenarbeit zwischen administrativen, forschungsunterstützenden und wissenschaftlichen Aufgaben entwickeln: „*Die traditionelle Arbeitsteilung zwischen Forschung und forschungsnahen bzw. forschungsunterstützenden Bereichen wird durch den digitalen Wandel zum Teil aufgelöst und durch neue Kooperationsformen und Aufgabenzuschnitte ersetzt.*“ (Rat für Informationsinfrastrukturen, 2019, S. 19). An diesen Stellen finden sich eine Reihe von Aufgaben, insbesondere im Bereich Forschungsdatenmanagement wieder, die sich an den Schwellen zwischen unterschiedlichen Stakeholder-Gruppen verorten lassen und die bislang keiner existierenden Berufsgruppe direkt zugeordnet werden können. Im Hinblick auf das Ziel einer transparenten Wissenschaft und dem Zugang zu offenen Forschungsdaten bleibt zu prüfen, ob an Forschungseinrichtungen tätige Data Stewards diese Schwellentätigkeiten übernehmen können und auf welche Weise dies geschehen kann.

⁴In dieser Arbeit wird sowohl der englischsprachige Begriff als auch die deutsche Übersetzung „Grenze“ verwendet. Ein Grund dafür ist die englischsprachige Herkunft der Theorien zu *Boundaries*. Außerdem impliziert der englische Begriff eine weniger scharfe Trennung von zwei Bereichen als der deutsche Begriff, sondern beschreibt auch sich überschneidende Bereiche. „Schwellen“ ist hierfür ein Wort, das dieser Implikation näher kommt.

1.2 Forschungsfragen

Aufbauend auf einer Sicht von Forschungsdatenmanagement als soziotechnisches System, mit einem Fokus auf soziale Aspekte, soll in dieser Arbeit das Konzept Data Stewardship näher betrachtet werden. Aufgrund einer bislang fehlenden adäquaten deutschen Übersetzung soll zunächst der Begriff Data Stewardship als Arbeitsgrundlage geschärft werden. Hierfür müssen die Begriffsentwicklung und die Bedeutungsdimensionen sowie das konzeptionelle Verständnis von Data Stewardship und Data Stewardship Tätigkeiten nachvollzogen werden. Daraus ergibt sich folgende erste Forschungsfrage:

1. Welche Bedeutung hat der englische Begriff Data Stewardship?

1.1: Wie hat sich der Gebrauch des Begriffs entwickelt?

1.2: Welche begrifflichen und konzeptionellen Bedeutungsdimensionen werden mit dem Begriff beschrieben?

Nachdem eine begriffliche Arbeitsgrundlage entwickelt wurde, soll im weiteren überprüft werden, ob sich Data Stewardship Tätigkeiten in die Schwellenbereiche zwischen unterschiedlichen beteiligten Akteursgruppen im Datenmanagement einordnen lassen. Hierzu muss untersucht werden, welche Tätigkeiten eine Zusammenarbeit erfordern und auf welche Weise diese Zusammenarbeit umgesetzt werden kann. Für diese Art von Fragestellungen lassen sich Untersuchungen zu sozialen Aspekten bei der Kollaboration unterschiedlicher Akteure, als theoretische Grundlage verwenden. Diesen Studien liegt das theoretische Konzept der Sozialen Welten, koordinierende Tätigkeiten an Schwellen zwischen sozialen Welten (*Boundary-Work*) und Kollaboration zwischen Sozialen Welten mit Hilfe von Artefakten (*Boundary Objects*) zugrunde. Es soll untersucht werden, inwieweit sich diese Grundlagen auf das Anwendungsfeld des Data Stewardships übertragen lassen. Demzufolge ergibt sich als zweite Forschungsfrage:

2. Lassen sich sozio-theoretische Modelle zu Kollaborationsprozessen zwischen Akteuren unterschiedlicher Sozialer Welten auf das Anwendungsfeld des Data Stewardship übertragen?

2.1 Für welche Tätigkeiten im Data Stewardship können für die Kollaboration hinderliche Grenzen identifiziert werden?

2.2 Welche Rolle nehmen Data Stewards bei der Zusammenarbeit ein?

2.3 Welche Methoden können hilfreich sein um Kollaborationsprozesse zu unterstützen?

Ziel dieser Arbeit ist ein ganzheitlicher Blick auf Data Stewardship einschließlich Aufgaben, Definitionen und Rollen. Es werden Vorschläge erarbeitet, Data Stewardship unter einem neuen Blickwinkel zu untersuchen. Es werden theoretische Grundlagen und mögliche Anwendungsfelder beleuchtet, die in weiteren Forschungsvorhaben empirisch getestet werden könnten. Ziel dieser Arbeit ist dabei die Erarbeitung eines Referenzmodells durch die Übertragung von Theorien zu Kollaborationen in soziotechnischen Systemen und des Agierens an Boundaries aus der Technik- und Wissenschaftssoziologie auf das Anwendungsfeld des

Data Stewardships.

Die Arbeit gliedert sich in folgende Kapitel:

Im Forschungsstand (Kapitel 2) wird in Kapitel 2.1 zunächst ein Überblick über neuere Entwicklungen im Bereich Forschungsdatenmanagement gegeben und dann in Kapitel 2.2 der theoretische Rahmen der Arbeit abgesteckt. In Kapitel 3 wird das methodische Vorgehen der Arbeit beschrieben. Daran schließt Kapitel 4 mit einer Definition des Data Stewardships an. Es werden in Kapitel 4.1 die Begriffs- und Bedeutungsentwicklung in der Literatur betrachtet, in Kapitel 4.2 konzeptionelle Unterschiede nachvollzogen und in Kapitel 4.3 ein eigener Arbeitsbegriff entwickelt.

In Kapitel 5 wird der theoretischen Rahmen dieser Arbeit näher beschrieben. Hierzu werden in Kapitel 5.1 Boundary-Work, in unterschiedlichen Formen und in der in dieser Arbeit verwendeten Form dargelegt und in Kapitel 5.2 ein Einblick in das im weiteren verwendete Kategorisierungsmodell von Boundaries gegeben. In Kapitel 5.3 wird das Konzept der Verwendung von Boundary Objects als Methode für Boundary-Work vorgestellt.

Die Einordnung von Data Stewardship in diesen theoretischen Hintergrund erfolgt in Kapitel 6. Hierzu werden zunächst in Kapitel 6.1 die Akteure im Data Stewardship, die in dieser Arbeit betrachtet werden beschrieben. Daran anschließend wird in Kapitel 6.2 eine Analyse von Data Stewardship-Tätigkeiten vor dem Hintergrund von Boundary-Work vorgenommen. Die Ergebnisse dieser Analyse werden in Kapitel 7 diskutiert. Dabei werden drei Thesen abgeleitet, die das Ergebnis dieser Arbeit darstellen. Kapitel 8 schließt mit einem Fazit und einem Ausblick auf weitere Forschungsfragen, die sich aus den Ergebnissen der Arbeit ergeben.

2 Forschungsstand

2.1 Neuere Entwicklungen im Forschungsdatenmanagement

Mit zunehmender Akzeptanz und steigendem Stellenwert des Forschungsdatenmanagements für einen offenen Wissenschaftsbetrieb im Sinne von *Open Science*⁵ lassen sich in den letzten Jahren eine Reihe koordinierender Aktivitäten auf nationaler und internationaler Ebene verfolgen, die eine Professionalisierung von Datenmanagementtätigkeiten über disziplinäre und institutionelle Grenzen hinweg voraussetzen. Um das Forschungsdatenmanagement von der institutionellen Ebene in eine breitere Koordination zu befördern, ist in Deutschland eine Nationale Forschungsdateninfrastruktur (NFDI) in Planung. Der Rat für Informationsinfrastrukturen (RfII) empfahl 2016 den Aufbau von koordinierten und kollaborativen Infrastrukturen für das Forschungsdatenmanagement sowie für die Personalentwicklung, um eine Fragmentierung von Angeboten in Deutschland zu verhindern und eine Vernetzung zu ermöglichen (Rat für Informationsinfrastrukturen, 2016). Diese Ziele, sowie ein Konzept für fachwissenschaftliche Konsortien innerhalb der NFDI, wurden 2017 weiter ausgeführt (Rat für Informationsinfrastrukturen, 2017a; Rat für Informationsinfrastrukturen, 2017b). 2018 wurde diese Vision einer „verteilte[n], vernetzte[n] Infrastruktur für Datenerzeuger und -nutzer, die durch verlässliche und nachhaltige Dienste-Portfolios generische und fachspezifische Bedarfe des Forschungsdatenmanagements in Deutschland abdeckt“ weiter beschrieben (Rat für Informationsinfrastrukturen, 2018b, S. 2; vgl. auch Rat für Informationsinfrastrukturen, 2018a). Die Ausschreibung für die erste Förderphase der NFDI-Konsortia ist am 6. Juni 2019 veröffentlicht worden.⁶ Mit dem Start der ersten NFDI-Konsortia kann davon ausgegangen werden, dass eine intensive Zusammenarbeit unterschiedlichster Akteure mit dem gemeinsamen Ziel einer vernetzten Forschungsdateninfrastruktur in einem noch höheren Maße als bisher gefordert werden wird.

Forschungsdaten sollen auch auf europäischer Ebene durch die European Open Science Cloud (EOSC) breit verfügbar sein. Die EOSC soll eine für Europa übergreifende Lösung für die Veröffentlichung und das Auffinden von Forschungsdatensätzen ermöglichen. Am 23. November 2018 wurde die EOSC offiziell eingeführt.⁷ Die Architektur der EOSC schafft eine übergreifende Infrastruktur, die bereits bestehende Forschungsdateninfrastrukturen vernetzt. Dabei müssen sowohl technische Interoperabilität zwischen unterschiedlichen Systemen, als auch koordinierende und kollaborative Aufgaben zwischen den „menschlichen“ Akteuren und in der Mensch-Technik-Interaktion gelöst werden. Ziele der EOSC sind u.a. die Schaffung einer gemeinsamen europäischen „*Data Culture*“ mit vernetzten, standardisierten und offen

⁵Unter dem Begriff Open Science (deutsch auch: Offene Wissenschaft) werden Strategien und Verfahren verstanden, die darauf abzielen, Forschung zugänglich, nach nutzbar und nachvollziehbar zu machen. Dabei gibt es unterschiedliche Kernkonzepte, zu denen *Open Research Data*, *Open Access*, *Open Metrics*, *Open Peer Review* aber auch *Citizen Science* und *Open Educational Resources* gehören. Grundprinzipien von Open Science sind Transparenz, Reproduzierbarkeit, Wiederverwendbarkeit und offene Kommunikation (AG Open Science, 2019).

⁶Siehe: https://www.dfg.de/foerderung/info_wissenschaft/2019/info_wissenschaft_19_37/index.html. Abgerufen am 21. Oktober 2019.

⁷Siehe: https://ec.europa.eu/info/news/commission-launches-european-open-science-cloud-2018-nov-23_en. Abgerufen am 21. Oktober 2019.

zugänglichen (wenn datenschutzrechtlich angebracht) Forschungsdaten in vertrauenswürdigen Repositorien, welche durch gut ausgebildetes Personal unterstützt werden (European Commission, 2018a). Die EOSC soll „*people, data, services and trainings, publications, projects and organisations across borders and scientific disciplines*“ miteinander verlinken (European Commission, 2018c, S.9). Hierfür sind folgende Aspekte besonders wichtig:

- **Open Science Policies**, die mit der EOSC übereinstimmen und Themen wie neue Formen der Impact-Messung, ein erweitertes Verständnis von guter wissenschaftlicher Praxis, Qualität und offene Formen der Kompetenzvermittlung einbeziehen.
- Die **FAIR Data Principles** (s.u.) als übergreifender Standard für die Vernetzung von Daten und Infrastrukturen.
- Die **Ausbildung von Kompetenzen und Fähigkeiten** für eine offene, datengetriebene Wissenschaft.
- Die Schaffung einer förderalen Infrastruktur, welche die offene Wissenschaft unterstützt und auf **vertrauenswürdigen Anbietern und Prozessen** beruht, z.B. durch Zertifizierungen von Repositories durch das *Core Trust Seal*⁸ oder eine *EOSC-ready* Zertifizierung für Software.

Die FAIR Data Principles (FAIR Data Prinzipien) haben sich als Rückgrat der EOSC herausgebildet. Sie wurden 2014 zusammen mit unterschiedlichen Akteuren aus Wissenschaft und Wirtschaft formuliert und 2016 durch eine Arbeitsgruppe der FORCE11 Community veröffentlicht, um das Datenmanagement und die wissenschaftliche Infrastruktur für eine ideale Auffindbarkeit und Nachnutzbarkeit von wissenschaftlichen Ergebnissen (Daten) zu gestalten (Wilkinson et al., 2016). Die Prinzipien sind dabei nicht als Standards anzusehen, sondern zeigen Ziele auf, welche die langfristige Auffindbarkeit und Nachnutzbarkeit von (Meta-)Daten ermöglichen sollen (Mons et al., 2017). „FAIR-konformes Datenmanagement“ zielt auf qualitativ hochwertige Forschungsergebnisse ab, welche sowohl von Menschen als auch insbesondere von Maschinen leicht auffindbar (*findable*), zugänglich (*accessible*), interoperabel (*interoperable*) und nachnutzbar (*reusable*) sind. Hervorgehoben wird die maschinelle Komponente bei der Ausführung der Auffindung, des Zugriffs, der Interoperabilität und der Nachnutzung. In Deutschland arbeitet insbesondere das Projekt GO FAIR⁹ mit den drei Säulen GO CHANGE (Policies, Anreize), GO BUILD (Technologien) und GO TRAIN (Aus- und Weiterbildung) an einer weiteren Implementierung der FAIR Prinzipien. Die FAIR Prinzipien sind folgendermaßen beschrieben:

- **Findability:** (Meta-)Daten sind sowohl für Maschinen als auch für Menschen leicht auffindbar.
 - F1 *(meta)data are assigned a globally unique and persistent identifier*
 - F2 *data are described with rich metadata (defined by R1 below)*
 - F3 *metadata clearly and explicitly include the identifier of the data it describes*
 - F4 *(meta)data are registered or indexed in a searchable resource*
- **Accessibility:** Die Zugänglichkeit der (Meta-)Daten ist durch (maschinenlesbare) Regelungen und Lizenzen klar geregelt.
 - A1 *(meta)data are retrievable by their identifier using a standardized communi-*

⁸Siehe: <https://www.coretrustseal.org/>. Abgerufen am 21. Oktober 2019.

⁹Siehe: <https://www.go-fair.org/>. Abgerufen am 21. Oktober 2019.

cations protocol

A1.1 *the protocol is open, free, and universally implementable*

A1.2 *the protocol allows for an authentication and authorization procedure, where necessary*

A2 *metadata are accessible, even when the data are no longer available*

- **Interoperability:** (Meta-)Daten sind standardisiert ausgezeichnet, so dass sie von Maschinen interpretiert und mit anderen Daten verknüpft werden können.

I1 *(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*

I2 *(meta)data use vocabularies that follow FAIR principles*

I3 *(meta)data include qualified references to other (meta)data*

- **Reusability:** Die Beschreibung der (Meta-)Daten und ihrer Eigenschaften erfolgt nach relevanten Standards, so dass sie einfach wiederverwendet und repliziert werden können.

R1 *meta(data) are richly described with a plurality of accurate and relevant attributes*

R1.1 *(meta)data are released with a clear and accessible data usage license*

R1.2 *(meta)data are associated with detailed provenance*

R1.3 *(meta)data meet domain-relevant community standards*

Jede Kategorie der FAIR Prinzipien ist jeweils in eine Reihe von Unterkategorien unterteilt, die spezifische Angaben für die Auffindbarkeit, den Zugang, die Interoperabilität und die Nachnutzbarkeit von Daten machen. Sie schlagen sich im Konzept der *FAIR Digital Objects* nieder, deren Kern Daten, Code oder Forschungsergebnisse in anderer Form ausmachen, die mit Persistent Identifiern (PIDs) versehen sind und in einem offenen, standardisierten Dateiformat dargestellt sind. Beschreibende Metadaten, deren Dokumentation ebenfalls bereit gestellt ist, sind dabei unerlässlich (Hodson et al., 2018).

FAIR und Open (im Sinne von Open Access/Open Data) werden als komplementäre Konzepte angesehen. FAIR-konforme Daten müssen nicht gleichzeitig offene Daten sein. Findable bedeutet, dass die Datensätze lokalisierbar sind, was durch eine ausführliche Beschreibung mit Metadaten realisiert wird. Accessible bedeutet, dass die Daten zugänglich sind, jedoch kann dieser Zugriff auch eingeschränkt sein (siehe Prinzip A1.2). Der Zugriff auf Daten soll demnach so offen wie im ethischen und rechtlichen Rahmen möglich erfolgen und muss gleichzeitig so geschlossen wie nötig bleiben. Eine Möglichkeit ist, sowohl FAIR als auch Open als Skalenwerte anzusehen, die sich bei aufsteigenden Niveaus immer weiter annähern (Collins et al., 2018). Die Umsetzung der FAIR Prinzipien aus der menschlichen Perspektive bedeutet vor allem die Operationalisierung des Teilens und der Nachnutzung von Daten im Rahmen rechtlicher Vorgaben, die Beachtung von Qualitätsstandards und Forschungsintegrität sowie von ethischen Aspekten von Forschungsdaten und die Förderung des Datenmanagements und der Datenpublikation durch neue Reputationsmechanismen (Boeckhout, Zielhuis und Bredenoord, 2018).

Stakeholder bei der Umsetzung der FAIR-Prinzipien sind neben den wissenschaftlichen Communities und den zugehörigen Forschungseinrichtungen auch Entscheidungsträger auf strategischer und administrativer Ebene, die im Englischen als *Policymaker* bezeichnet wer-

den. Dazu gehören wissenschaftspolitische Einrichtungen, die Leitlinien zum Umgang mit Daten verabschieden. Außerdem sind auch Verlage, Infrastruktureinrichtungen, koordinierende Organisationen (wie die Research Data Alliance (RDA) oder GO FAIR), Forschungsförderer und Zertifizierungseinrichtungen (z.B. für Repositorien oder Standards) an der Umsetzung beteiligt. Nicht zuletzt werden Datenspezialist*innen aus den wissenschaftlichen Communities und/oder Forschungsinfrastrukturen, wie z.B. Bibliotheken als Stakeholder angesehen (Hodson et al., 2018). Für die Umsetzung der FAIR Prinzipien wurden im *FAIR Data Action Plan* folgende Akteure identifiziert (ebd., S. 3):

- **Forschungsgemeinschaften:** Forschende aus allen Forschungsbereichen, die sich um Disziplinen, Datentypen oder wissenschaftliche Fragestellungen herum gruppieren.
- **Datenservices:** Fachspezifische oder institutionelle Repositorien, (digitale) Forschungsinfrastrukturen sowie offene und kommerzielle Anwendungen und Services für das wissenschaftliche Arbeiten.
- **Data Stewards:** Support-Mitarbeiter*innen aus den wissenschaftlichen Communities und Bibliotheken sowie administratives Personal für Repositorien.
- **Normungsgremien:** Offizielle Einrichtungen und Konsortien, die Standardisierungsverfahren, z.B. für die Zertifizierung von Repositorien oder die Akkreditierung von Curricula, koordinieren.
- **Globale Koordinationsinitiativen:** RDA, das Committee on Data (CODATA), FORCE11, GO FAIR und andere.
- **Entscheidungsträger:** Stellen, an denen Leitlinien zum Umgang mit Daten definiert werden, z.B. Regierungen, internationale Vereinigungen wie die Organisation for Economic Co-operation and Development (OECD), Forschungsförderer, Institutionen oder Verlage.
- **Forschungsförderer:** Die Europäische Kommission, nationale Forschungsförderer (z.B. die DFG), Stiftungen sowie andere Einrichtungen zur Forschungsförderung.
- **Institutionen:** Universitäten und Forschungseinrichtungen sowie sonstige Einrichtungen an denen Forschung praktiziert wird.
- **Verlage:** Kommerzielle oder non-profit Verlage, die in Open Access oder Subskriptionsverfahren wissenschaftliche Publikationen und Daten veröffentlichen.

Die Europäische Kommission empfiehlt für die Umsetzung der FAIR Data Prinzipien, neue Arten von Kompetenz- und Tätigkeitsfeldern zu entwickeln, die sich in der Entwicklung von konkreten Professionen abbilden sollen. Für eine Daten-getriebene Forschung werden speziell ausgebildete *Data Scientists* benötigt, die Projekte aus unterschiedlichen Fachgebieten begleiten und in Kooperation mit Fachwissenschaftler*innen Datenanalysen und (statistische) Auswertungen durchführen. Gleichzeitig betont die Europäische Kommission die Notwendigkeit der Ausbildung von Data Stewards, die während des Forschungsprozesses und darüber hinaus im Hinblick auf die spätere Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit der im Forschungsprozess entstehenden Daten das Datenmanagement sicherstellen und betreuen: „*Steps need to be taken to develop two cohorts of professionals to support FAIR data: data scientists embedded in those research projects which need them, and data stewards who will ensure the management and curation of FAIR data.*“ (Hodson et al., 2018, S. 10).

Eine Übersicht über benötigte Kompetenzen im Zuge der Einführung der EOSC wurden im Projekt im Workpackage 7 „Skills“ erarbeitet.¹⁰ Hintergrund hierfür sind für die EOSC formulierten Anforderungen an Fähigkeiten und Kompetenzen im Bereich Data Stewardship und Data Science. Da bisher jedoch klare Definitionen und Einigkeit über Kompetenzen und Rollenverteilungen in diesem Bereich fehlen, wurde im EOSCpilot ein „Kompetenzframework“ erarbeitet, welches auf bereits bestehenden Frameworks aufbaut¹¹ und diese insbesondere im Hinblick auf FAIR Data und Open Science Praktiken ergänzt. Data Stewardship wird als eine Aufgabe angesehen, die die Bereiche Datenmanagement und -kuration, Data Science und Data Analytics, Data Service Engineering und Fachwissenschaft an ihren Überschneidungsgrenzen verbindet. Im Framework werden Fähigkeiten in drei Level (verstehen - anwenden - evaluieren) unterteilt. Diese Level werden den Rollen der Forschenden, der „Data Scientists/Analysts“, der „Data Service Engineers“ und der „Data Manager/Curator“ zugeordnet. Das Framework unterscheidet außerdem zwischen Service Nutzer*innen und -Anbieter*innen. Insgesamt werden 59 Fähigkeiten definiert, die folgende Kompetenzbereiche abdecken:

- Planen und entwickeln (Plan and design)
- Erheben und prozessieren (Capture and process)
- Integrieren und analysieren (Integrate and analyse)
- Bewerten und bewahren (Appraise and preserve)
- Publizieren und veröffentlichen (Publish and release)
- Zugänglich machen und auffinden (Expose and discover)
- Steuern und beurteilen (Govern and assess)
- Reichweite und Ressourcen (Scope and resource)
- Beraten und Befähigen (Advise and enable)

Vorreiter in der Formalisierung und der Anwendung eines solchen Verständnisses von Data Stewardship sind die Niederlande, insbesondere die TU Delft und das Dutch Techcenter for Lifesciences (DTL). An der TU Delft wird seit Beginn 2018 ein Data Stewardship Programm durchgeführt, das bis 2020 gefördert wird. Data Stewards stehen für jede Fakultät zur Verfügung und werden durch die Bibliothek koordiniert. Die Data Stewards sind dabei in allen Daten-relevanten Bereichen der Universität vernetzt und fungieren als erste Ansprechpartner für alle (fachspezifischen) Aktivitäten zum Datenmanagement. Sie agieren unterstützend für die Forschenden und verweisen sie, wenn notwendig, an jeweils verantwortliche Stellen. Zudem geben sie Trainings in Form von Workshops und helfen bei der Erstellung von Datenmanagementplänen (DMPs), bei Fragen zu Policies von Zeitschriften und Forschungsförderern und bei der Veröffentlichung und Archivierung von Forschungsdaten. Gleichzeitig sind sie Multiplikatoren für eine gute wissenschaftliche Praxis im Sinne von Open Science und wirken auch an der Gestaltung von institutionellen Policies zum Umgang mit Forschungsdaten mit (Teperek, 2018).

¹⁰Siehe: <https://eoscpilot.eu/themes/skills>. Abgerufen am 21. Oktober 2019.

¹¹Zum Beispiel das EDISON Data Science Competence Framework, siehe: <http://doi.org/10.5281/zenodo.1044346>

2.2 Theoretische Grundlagen

Als theoretische Grundlage in dieser Arbeit wird das Konzept von Sozialen Welten und Grenzen sowie Interaktionsmöglichkeiten zwischen ihnen angewendet. In der Wissenschafts- und Technikforschung (engl: *Science and Technology Studies* (STS)) wird das *Social World Framework* als „*theory/methods package*“, also einer Verbindung von Theorien mit konkreten methodischen Forschungsansätzen verwendet (Clarke und Star, 2008). Dabei werden Beziehungen und Interaktionen zwischen Akteuren und Dingen im Sinne der Akteur-Netzwerk-Theorie (z.B. in Latour, 1996) untersucht. Hierzu gehören Arten von Repräsentationen, Prozessen und Arbeitsabläufen und Diskursgeflechten. Das Framework bildet dabei Kontextualitäten und situative Bedingungen, Fluidität und vergangene, sowie zukünftige Veränderungen in den Interaktionen zwischen und innerhalb der untersuchten Gruppen ab. In diesem Ansatz charakterisieren besonders gemeinschaftliche Handlungen, sowie die interaktive Konstruktion von Sinn und Bedeutung das Konzept der Soziale Welten (Garrety und Badham, 2000).

Das Konzept entstammt der amerikanisch geprägten soziologischen Tradition des symbolischen Interaktionismus, dessen Grundlagen in der Chicagoer Schule liegen. Im symbolischen Interaktionismus wird davon ausgegangen, dass die „Wirklichkeit“ und die Gesellschaft durch andauernde Interaktionsprozesse unter Akteuren sowie zwischen Akteuren und Gesellschaft hergestellt wird. Diese Interaktionen bedingen sich gegenseitig und haben die Konstruktion der Wirklichkeit zur Folge. Es werden daher insbesondere Relationen und Interaktionen zwischen den einzelnen Akteuren untereinander und mit der Gesellschaft betrachtet. Besonders von Interesse sind dabei Beziehungen zwischen subjektiven Identitätsentwürfen. Es wird davon ausgegangen, dass Identität sowohl durch die eigene Erfahrungswelt beeinflusst wird, als auch gleichzeitig diese Erfahrungswelt konstruiert. Diese reziproke Beziehung haben dynamische Prozesse von Sinnbildung, Ausbildung von Rollen, Werten und Normen innerhalb der Gesellschaft zur Folge (Reiger, 2009).

Strauss (1978) bezeichnet Soziale Welten auch als „*universes of discourse*“, die sich durch identitätsbildende Aktivitäten, Technologien und Organisationen (im Sinne von Zusammenarbeit, nicht unbedingt im institutionalisierten Sinne) auszeichnen. Da Sinnbildungsmechanismen innerhalb der Sozialen Welten durchgehend aktiv sind, gelten Prozesshaftigkeit und Veränderlichkeit als weitere Merkmale. Zugehörigkeitsgefühle und Ausschlüsse oder Abgrenzungen innerhalb Sozialer Welten sind immer eng verbunden mit Verhandlungen um Authentizitäts- und Machtdebatten (Strauss, 1978). Welche Weltsicht als konstituierend angesehen wird und wo die Grenzen zwischen unterschiedlichen Welten verlaufen ist somit eine konstante Verhandlungsgrundlage für Sinngebungsverfahren, sowohl subjektiv für einzelne Akteure als auch für gesellschaftliche Zusammenhänge. Interaktionen zwischen Akteuren innerhalb einer Sozialen Welt drücken sich in Diskursen aus, in denen die Entwicklung eines gemeinsamen Modus Operandi angestrebt wird (Clarke und Star, 2008). Im *Practical Turn* in der Soziologie wird davon ausgegangen, dass sich Wissen in Praktiken widerspiegelt. Das soziale Umfeld (oder die Soziale Welt) der Akteure wird durch ihre Praktiken (oder Tätigkeiten) strukturiert und fungiert gleichzeitig als (begrenzender) Rahmen für diese Tätigkeiten. Tätigkeiten werden auf Grundlage von gemeinsamen Interessen, Werten und ihren sozialen Positionen von einzelnen Akteure in ihrer Sozialen Welt ausgeführt (Reckwitz, 2003); Bongaerts, 2007).

Dabei ist (gruppenspezifisches) Wissen in diese Praktiken eingebettet und kann an Stellen, wo es zu Kollaboration zwischen den Akteursgruppen kommen muss, zu (kognitiven) „Grenzen“ - englisch „*boundaries*“ - oder Differenzen (Knowledge Boundaries) führen (Carlile, 2002). Grundlage hierfür kann beispielsweise das Vorliegen von „explizitem“, d.h. artikulierbarem Wissen und „implizitem“ Wissen, das sich in z.B. in Interaktionen oder Routinen widerspiegelt, sein (vgl. hierzu Polanyi und Nye, 2015). Hsiao, Tsai und Lee bezeichnen Boundaries als „*a demarcation, or a sphere of activities, that marks the limits of an area, which may include knowledge, tasks, as well as hierarchical, physical, geographical, social, cognitive, relational, cultural, temporal/spatial, divisional, occupational, and disciplinary boundaries*“ (Hsiao, Tsai und Lee, 2012, S. 1). Aus dieser Fülle von Aspekten wurden für die Untersuchung von Boundaries in den Sozialwissenschaften bislang besonders Grenzen zwischen sozialen und kollektiven Identitäten, zwischen Klassen, Gender und Ethnien, zwischen Wissen, Wissenschaft und Berufsständen und Grenzen zwischen Communities, Nationalitäten oder Nationen aufgegriffen (Lamont und Molnár, 2002).

Boundaries charakterisieren sich durch ihre absolute Ambiguität. Sie existieren zwischen zwei Sozialen Welten und gehören sowohl zu beiden Welten, stellen aber auch gleichzeitig eine Art Niemandsland dar, da sie weder zu der einen noch zu der anderen Welt eindeutig zugeordnet werden können. Akkerman und Bakker beschreiben diese Ambiguität als „*both-and*“- und „*neither-nor*“- Phänomen (deutsch: sowohl-als-auch- und weder-noch-Phänomen) (Akkerman und Bakker, 2011, S. 19). Je nach Situation ist es sinnvoll, Boundaries zu festigen oder ihre Durchlässigkeit zu erhöhen. Gefestigte Boundaries können Akteure in Sozialen Welten bei einer Identitätsbildung unterstützen, indem zwischen ihnen und anderen Akteuren jenseits der Grenze ihrer Welt unterschieden wird (*othering*). Beispielsweise ist die Unterscheidung zwischen Wissenschaft und „Nicht-Wissenschaft“ und die klare Grenzziehung in Teilen der Wissenschaftskommunikation, eine Form von Boundary-Festigung (Gieryn, 1983.). Boundaries, die sich bei Kollaborationen als hinderlich erweisen, können auch als Startpunkt von Diskursen und Verhandlungen fungieren und dabei zu einem Lerneffekt für alle beteiligten Akteure beitragen (Akkerman und Bakker, 2011). Diese andere Form des Umgangs mit Boundaries, das *boundary-crossing*, bzw. *boundary-spanning* ist Grundlage für die Betrachtung des Wirkens an Grenzen (*Boundary-Work*) in dieser Arbeit.

Rezazade M. und Hawkins (2012) unterscheiden zwischen *Boundary Spannern*, *Boundary Objects*, *Boundary Practices* und *Boundary Discourse*. Boundary Spanner sind „*human agents who translate and frame information from one community to another in an effort to promote coordination*“ (Rezazade M. und Hawkins, 2012, S. 1803). Diese Menschen handeln somit zwischen Akteursgruppen, um durch übersetzende Tätigkeiten (kognitive) Grenzen zu überwinden und so Kollaborationen zu befähigen. Boundary Objects als materielle oder konzeptionelle Artefakte können dazu beitragen, Wissen zu de-kontextualisieren und de-personalisieren, so dass eine Transformation von Wissen (und Praktiken) über Grenzen hinweg bewirkt werden kann. Boundary Practice ist eine besondere Form des Boundary Spannings, bei der unterschiedliche Akteure an Grenzen in kollektiven Aktivitäten aufeinander treffen. Boundary Discourse bezeichnet den Inhalt der Kommunikation, die an Boundaries zwischen den unterschiedlichen Akteuren stattfindet. Boundary Spanning Funktionen und Konzepte wurden vor allen in der Organisations- und Managementtheorie mit Bezug auf Kollaboration, Innovation und Produktion untersucht (u.a. in Tushman und Scanlan, 1981;

Fleming und Waguespack, 2007; Williams, 2013).

Für diese Arbeit sind insbesondere Aktivitäten des Boundary Spannings mit Hilfe von Boundary Objects (BOs) relevant. Das Konzept der Boundary Objects (deutsch: Grenzobjekte) wurde 1988 von Leigh Star eingeführt und 1989 von Star und Griesemer in ihrer oft zitierten Studie zur Entstehung des Berkley Museum of Vertebrate Zoology und der wissenschaftlichen Kooperation des dort arbeitenden Personals weiter entwickelt (Star und Griesemer, 1989). Darin untersuchten sie die Zusammenarbeit zwischen Fachwissenschaftler*innen (z.B. Biolog*innen) und „Amateuren“, also z.B. Jäger*innen oder „Trapper“ und deren Kommunikation und „Übersetzung“ von Begrifflichkeiten und Konzepten untereinander. Sie fanden heraus, dass zwei Faktoren zu der erfolgreichen Entwicklung des Museums beigetragen hatten: Methodische Standardisierungsverfahren und die Entwicklung von Boundary Objects. Boundary Objects werden folgendermaßen definiert:

„Boundary objects are objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. They may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is key in developing and maintaining coherence across intersecting social worlds.“ (Star und Griesemer, 1989, S. 393)

Als Boundary Objects können also nicht nur physische Artefakte, sondern auch abstrakte Konzepte in Frage kommen. Wie im obigen Beispiel deutlich wurde kommen BOs an den Schwellen zwischen unterschiedlichen Sozialen Welten vor. Sie werden häufig auch an den Grenzen unterschiedlicher *Communities of Practice* (CoP), also Gruppen von Personen, die durch Wissensbereiche, Tätigkeiten und Ziele eine Gemeinschaft bilden, verortet (Wenger, 2000). Sie dienen als Übersetzungswerkzeuge (*translation devices*) und helfen dabei, Verständnis, Kommunikation und Zusammenarbeit über unterschiedliche CoPs und auch unterschiedliche Soziale Welten hinweg zu ermöglichen. Dies geschieht dadurch, dass BOs flexibel genug sind um für unterschiedliche Welten gemeinsame Informationsbedürfnisse zu erfüllen, jedoch ihre genaue Bedeutung innerhalb einzelner Sozialen Welten zu erhalten, ohne einen Bedeutungsverlust zu erfahren (Bowker und Star, 1999, S. 297).

Huvila et al. (2017) geben einen umfassenden Überblick über die Verwendung von BOs in den Informationswissenschaften. Sie zeigen u.a. unterschiedliche Konzepte auf, die im Bezug auf BOs erarbeitet wurden. Dabei handelt es sich um Aktivitäten (wie *boundary breaking*, *boundary spanning*, *boundary work*), Dinge (z.B. *boundary concepts*, *boundary organizations*) oder Betrachtungen zu unterschiedlichen Arten von Boundaries (z.B. *evolutionary boundaries*, *knowledge boundaries* etc.) (Huvila et al., 2017, S. 7). Zudem zeigen sie, wie sich in der Literatur der Fokus in den letzten Jahren von der Operationalisierung von BOs als einfachen Übersetzungswerkzeugen hin zur Analyse von sozialen Relationen weiterentwickelt hat. Sie folgern, dass in den Informationswissenschaften das Konzept von BOs dabei helfen kann, sozio-materielle Aktivitäten in Relation zu Informationen aufzu-

zeigen. Dabei stehen insbesondere die Verknüpfungen zwischen Artefakten („Objects“), ihren Charakterisierungen und Rollen („*objects in relation to practices*“), Praktiken im Bezug auf die Artefakte („*objects as outcomes of practices*“) und zugehörige Verwendungen („*practices in relation to objects*“) im Vordergrund. Aus epistemologischer Sicht können BOs in den Informationswissenschaften dazu verwendet werden, Theorien und Konzepte der Fachdisziplin („*conceptualizations of objects, practices and interactions of objects and practices*“) zu betrachten, zu kritisieren und neu zu definieren (ebd., S. 17, Tabelle 4).

Untersuchungen zu Stewardship-Aktivitäten vor diesem Hintergrund gibt es bislang noch nicht. Boundaries wurden vor allem in den Informationswissenschaften bisher im Bezug auf Dokumente, Standards, Informationssysteme und standardisierte Vokabulare betrachtet (Huvila et al., 2017). Studien zu soziologischen Aspekten bei der Kollaboration im Forschungsdatenmanagement, zu Problemen, Grenzen und Methoden der gemeinsamen Verständigung sind bislang kaum vorhanden. Zuiderwijk und Janssen haben die Veröffentlichung und Nachnutzung von Open Data (hierbei eher im Sinne von Open Government Data als von Open Research Data) und deren Hindernisse unter sozio-technologischen Aspekten betrachtet (Zuiderwijk und Janssen, 2014). 2016-2017 untersuchte das Projekt *Data Curator: Who is s/he?* Rollen, Wahrnehmungen und Aufgaben von Datenkurator*innen in einem interdisziplinären Kontext (Tammaro et al., 2017; Tammaro und Casarosa, 2018). Ein Ergebnis dieser Studie ist, dass dieser Ansatz nicht nur pragmatische und technische Faktoren, sondern auch sozio-technologische Ansichten des Berufsbilds liefern kann. Gregory et al. haben vor kurzem die Suche nach Daten (*data search*) erstmals aus einer sozio-technologischen Perspektive untersucht und im Hinblick auf ein verbessertes Systemdesign ausgewertet. Sie folgerten, dass Nutzer*innen von Daten nicht als Forschende in begrenzten wissenschaftlichen Disziplinen mit festgelegten Forschungspraktiken agieren, sondern vielmehr als dynamische Communities, denen Forschungsinteressen oder -techniken zugrunde liegen und in denen Daten als „Schaltbrett“ für Kollaborationen und Kreativität verwendet werden: „*Researchers form new collaborations in order to share, access and make sense of data. These collaborations help to grow professional networks and can inspire new approaches to future projects.*“ (Gregory et al., 2019, S. 14).

In welcher Form Data Stewards mit diesen Communities interagieren können und welche Kollaborationsmechanismen angewendet werden blieb jedoch offen. Auch offen blieb bisher, ob sich bestehende theoretische Modelle aus den oben beschriebenen Kontexten von Interaktionen und Kollaborationen zwischen Sozialen Welten und über Boundaries hinweg für die Untersuchung von Stewardship Aktivitäten eignen können. Dies liegt zu einem großen Teil auch daran, dass Data Stewardship gerade im Bezug auf die Entwicklung der FAIR-Prinzipien als Schlüsselaufgabe für das Datenmanagement betrachtet wurde, ohne dass konkrete operative Funktionen, Rollen, Tätigkeiten und Arbeitsteilungen abschließend geklärt sind. Die Übertragung theoretischer Modelle zu Kollaborationsmechanismen vor dem Hintergrund des Social World Frameworks kann hierfür potentiell nützlich sein und wird in dieser Arbeit erprobt.

3 Methode

Um den Begriff Stewardship bzw. Data Stewardship zu erfassen, Begriffsdefinitionen nachzuvollziehen und Veränderungen in der Interpretation des Begriffs zu verfolgen wird in dieser Arbeit zunächst eine Literaturanalyse vorgenommen. Für die Analyse wird eine umfassende Recherche in der Fachliteratur zu den Themengebieten (Forschungs-)Datenmanagement, Data Curation und Data Preservation durchgeführt und relevante Publikationen wurden auf eine Definition oder eine Erklärung von „Stewardship“ überprüft. Dabei werden Definitionsansätze aus dem Bereich der Wirtschaft und Unternehmensführung zum Thema Data Stewardship und Governance bewusst weitestgehend ausgeklammert, da der Fokus der vorliegenden Arbeit auf Data Stewardship in der Wissenschaft und im Bezug zu Forschungsdaten liegt. Für die Begriffsdefinition werden nach der Auswertung der Publikationen einzelne Beispiele herausgegriffen und miteinander verglichen. Dabei werden induktiv thematische Schwerpunkte des Begriffs identifiziert und ein Verständnis von Data Stewardship als Grundlage dieser Arbeit entwickelt.

Um die Übertragbarkeit bestehender Theorien zu disziplinübergreifender Kollaboration und Zusammenarbeit zwischen Sozialen Welten zu prüfen wird zunächst der theoretische konzeptionelle Rahmen gelegt. Es werden Formen von Schwellentätigkeiten oder Grenzarbeit, als Boundary-Work bezeichnet, und die Verwendung von Boundary Objects als Werkzeuge für Boundary-Work anhand der Grundlagenliteratur näher beschrieben. Boundary-Work ist in dieser Arbeit nicht mit Abgrenzungsaktivitäten gleichzusetzen, wie in der von Gieryn (1983) entwickelten Definition, sondern mit Aktivitäten und Hilfsmitteln, die bei interdisziplinärer Zusammenarbeit helfen sollen. Es wird hierfür der Begriff *Boundary Spanning* in Sinne einer Überbrückungstätigkeit über bestehende Grenzen hinweg verwendet. Carlile (2004) hat unterschiedliche Arten von Knowledge Boundaries in syntaktische, semantische und pragmatische Boundaries kategorisiert und entsprechende Aktivitäten für die Grenzüberbrückung beschrieben. Diese werden als theoretische Grundlage verwendet.

Dem *Practical Turn* in der Soziologie folgend wird davon ausgegangen, dass sich Wissen in Praktiken widerspiegelt, die sich in die Sozialen Welten unterschiedlicher Akteursgruppen einordnen lassen. Für diese Arbeit werden auf Grundlage einer angenommenen situativen Verordnung des institutionellen Data Stewardships die Akteure Infrastruktureinrichtungen, Entscheidungsträger und Forschende ausgewählt und ihre Tätigkeiten im Data Stewardship näher betrachtet. Forschungsdatenmanagement betrifft nicht nur diese drei Akteursgruppen sondern auch eine Reihe anderer Stakeholder, z.B. Zertifizierungsorganisationen oder Verlage und Fachzeitschriften (Steinhof, 2018). Die vorliegende Auswahl der betrachteten Akteursgruppen beruht auf Darstellungen von Datenmanagement- und Stewardship-Aktivitäten, die zwischen forschungsnahen, infrastrukturellen und administrativen Dimensionen unterscheiden (Verheul et al., 2019, siehe Abbildung 3 ; Scholtens et al., 2019b, siehe Abbildung 2).

Auf Grundlage des EU-Projektes EOSCpilot können 59 Tätigkeiten im Bereich des Data Stewardship identifiziert werden. Für diese Tätigkeiten wird der Grad der Beteiligung der unterschiedlichen Akteure untersucht. Das Projekt-Deliverable D 7.3 *Skills and capability framework* (Whyte et al., 2018) beschreibt diese Tätigkeiten in unterschiedliche Abschnitte

gegliedert jeweils für Nutzende von Datenservices und Anbieter*innen ebendieser. In der vorliegenden Arbeit wird diese Unterteilung nicht verfolgt. Vielmehr werden die Tätigkeiten, wenn zutreffend, den jeweiligen Akteursgruppen zugeordnet. Dafür werden Kreisdiagramme verwendet, die in Anhang A eingesehen werden können. Die Kreise symbolisieren dabei die Sozialen Welten der Akteure (Infrastruktureinrichtungen = rot, Entscheidungsträger = grün, Forschende = blau). Der gelbe Kreis bezeichnet die Tätigkeit und ihre Verortung innerhalb der Sozialen Welt und relativ zu den anderen Akteuren. Die Ergebnisse sind im Text tabellarisch dargestellt. Wenn eine Akteursgruppe an einer Tätigkeit beteiligt ist, wird dies farblich gekennzeichnet, wobei die Intensität der Beteiligungen nicht abgebildet wird. Die Begründung für die Zuordnung der Tätigkeiten zu den einzelnen Akteursgruppen erfolgt im Text. Es handelt sich hierbei um grundsätzliche Überlegungen, die so sicherlich nicht vollständig in jeder Forschungseinrichtung und ihren Organisationsstrukturen abgebildet werden. Die Analyse kann jedoch als Rahmen angesehen werden um Strukturen und Schnittstellen zwischen den einzelnen Akteursgruppen festzustellen.

Die Analyse ermöglicht eine Identifikation der Tätigkeiten im Data Stewardship, an denen mehrere Akteure beteiligt sind. Für diese Tätigkeiten lassen sich potentiell auftauchende Boundaries beschreiben, die den von Carlile definierten Grenzarten zugeordnet werden können. Es wird, Carliles Verständnis von Zusammenarbeit zwischen unterschiedlichen Akteursgruppen folgend, davon ausgegangen, dass es bei solchen Tätigkeiten einer Überbrückung von Knowledge Boundaries bedarf. Entsprechende Überbrückungsaktivitäten, die als Übertragung (Transfer), Übersetzung (Translate) und Transformation (Transform) bezeichnet werden, lassen sich den jeweiligen Tätigkeiten zuordnen.

Ziel dieser Arbeit ist, das Konzept des Data Stewardships und die Tätigkeiten von Data Stewards in den Rahmen des sozio-theoretischer Ansatzes von Boundary-Work innerhalb kollaborativer Tätigkeiten einzuordnen und Erklärungs- sowie Lösungsansätze für dabei entstehende Herausforderungen herauszuarbeiten. Es wird erprobt, die bestehenden theoretischen Grundlagen auf das Anwendungsfeld des Data Stewardships zu übertragen. Hierfür wird die Analyse abschließend mit einem Fokus auf vorkommende Grenzarten und entsprechende Aktivitäten unter Einbezug der theoretischen Grundlagen des Boundary Spannings und der Verwendung von Boundary Objects diskutiert. Es wird herausgearbeitet, welche Rolle Data Stewards bei einer Zusammenarbeit im Data Stewardship einnehmen können und welche Tätigkeitsbereiche und Maßnahmen anfallen können. Zudem wird diskutiert, welche bereits bestehenden Artefakte als Boundary Objects von Data Stewards verwendet werden könnten.

4 Data Stewardship

4.1 Begriff und Bedeutungsentwicklung

Der Begriff Stewardship stammt aus dem Englischen und wird im Deutschen mit *Verwaltung* übersetzt.¹² Jedoch wird Stewardship in vielen Fällen als Anglizismus verwendet, vor allem in den Bereichen Business und Management. Hier bezeichnet Stewardship die Verantwortlichkeit für ein bestimmtes Produkt (z.B. *Product Stewardship*)¹³ oder einen Management-Prozess (z.B. *Asset Management*)¹⁴. Im Englischen bedeutet Stewardship „*the conducting, supervising, or managing of something*“ mit einem Fokus auf „*careful and responsible management of something entrusted to one's care*“ (Mariam-Webster.com, 2019), also die aktive Verantwortungsübernahme für etwas. Gebräuchlich ist der Begriff im englischsprachigen Raum auch im Bereich der Ökologie für den verantwortungsvollen Umgang mit natürlichen Ressourcen. Hierbei wird insbesondere die Verantwortungsübernahme für andere (im Sinne der breiteren Gemeinschaft) sowie ethische und moralische Aspekte im Hinblick auf nachfolgende Generationen (*future generations*) betont (z.B. in Worrell und Appleby, 2000). Ökologische Stewardship bezieht dementsprechend den Wert und die Endlichkeit von natürlichen Ressourcen, sowie deren Allmende-Implikation mit ein.

Neben der Ökologie wird der Begriff Stewardship in Form der Stewardship-Theorie in der Management-Theorie und der Unternehmensführung gebraucht. Diese Theorie bezeichnet nach Donaldson und Davis (1991) einen Führungsansatz, dessen Merkmale, im Gegensatz zur Principal-Agent-Theorie nach Berle und Means (z.B. in Berle und Means, 1991), geprägt sind von Verantwortungsübernahme und Akteuren, die in der Unternehmensführung als „gute Verwalter“ für ein kollektives Ziel (z.B. die Reputation des Unternehmens) arbeiten. Der Führungsstil ist langfristig orientiert und beruht auf einer klaren Teamorientierung sowie auf dem Schaffen von Vertrauen zwischen den Teammitgliedern und anderen Beteiligten (Velte, 2010).

Obwohl der Begriff, gerade in Bezug auf den Umgang mit Daten, in den letzten Jahren sehr häufig verwendet wurde, sind Definitionen in der Literatur bislang nicht einschlägig oder konsistent. Stewardship in Bezug auf Daten wurde erstmals 1991 in der Directive 8320.1 „*DoD Data Administration*“ des amerikanischen Department of Defense verwendet. Daten können „*under the stewardship*“ sein, was als die Verantwortung für das Management der Daten ohne dass ein Besitzanspruch vorliegt beschrieben wird (Department of Defense, 1998).

1996 beschrieb Dawes Information Stewardship im Kontext des Umgangs mit Informationen der öffentlichen Verwaltung (*government information*):

„*A stewardship principle views agencies as stewards of government information,*

¹²Siehe: <https://de.langenscheidt.com/englisch-deutsch/stewardship>. Abgerufen am 21. Oktober 2019.

¹³Siehe: <https://de.eurochemagro.com/nachhaltigkeit/product-stewardship/>. Abgerufen am 21. Oktober 2019.

¹⁴Siehe: <https://www.global.assetmanagement.hsbc.com/about-us/responsible-investing/stewardship>. Abgerufen am 21. Oktober 2019.

not owners. Stewardship focuses on the accuracy, integrity, and preservation of information holdings. It does not fix a single point of responsibility for such issues as accuracy, validity, security, use, description, or preservation. Instead, stewardship conveys the idea that every government organization is responsible for handling information with care and integrity, regardless of its original source. It demands that government information be acquired, used, and cared for as a jurisdiction-wide resource." (Dawes, 1996, S. 393)

Neben der Verwendung des Begriffs im Umgang mit wissenschaftlichen Daten, bezieht sich Data Stewardship in direkter Verbindung zu Data Governance vornehmlich auf das Management von Daten in der Wirtschaft. Mit der Entwicklung von *Data Warehouses*¹⁵ wurde die Rollenverteilung im Datenmanagement genauer definiert. Nach wie vor bezieht sich die Verwendung des Begriffs Steward(ship) auf den Unterschied zum „ownership“ von Daten und bezeichnet die Verantwortlichkeit für Daten trotz des fehlenden Besitzanspruchs. Data Stewardship Teams wird die Verantwortlichkeit für das Datenmanagement übertragen. Damit können eventuell auftretende Autoritätskonflikte, die sich auf Recht für die Datenverwaltung innerhalb von Unternehmen oder Teams beziehen, umgangen werden:

„The people who manage data processes are, in the current data warehouse vernacular, called data stewards. [...] [A] data steward is a person who manages information and those activities that encompass data creation, capture, maintenance, decisions, reporting, distribution, and deletion. Therefore if a person performs any of these functions on a set of data, he or she is a data steward. [...] In the definition of steward, there is a caption to emphasize: 'One who manages another's property ...' Many times project managers have complained they can not move their project past a certain point because the stakeholders could not agree on who 'owned' the data. This is dead center a stewardship issue. No steward owns the data. The data is owned by the organization just as surely as the organization owns its name, trademarks, cash, and purchased equipment. The debate on ownership is not really about ownership, but usually centers on who has the authority to approve a change to the data. The answer is the data stewardship team." (Dravis, 2004, S.36)

Plotkin (2014) beschreibt Data Stewardship in Unternehmen als die Ausführung von Data Governance. Data Governance bedeutet in diesem Sinne ein System von Entscheidungen zu Rechten und Verantwortungen für alle Prozesse des Datenmanagements. „Governed data“ in Unternehmen ist vertrauenswürdig (*trusted*), verständlich und verstanden (*understood*) und betreut von verantwortlichen (*accountable*) Personen, den Data Stewards. Data Stewardship ist die Formalisierung dieser Verantwortlichkeit durch entsprechende Personen im Unternehmen. Data Governance und Data Stewardship beruhen auf Policies, Prozessen und Verfahren, die den genauen Rahmen und die Abläufe des Datenmanagements vorgeben. Plotkin unterscheidet in seiner Definition zudem zwischen *Business Data Stewards* und *Technical Data Stewards* in Unternehmen. Business Data Stewards übernehmen die Verantwortung

¹⁵Ein Data Warehouse ist eine zentrale Datenbank, in der Daten aus heterogenen Quellen zusammengeführt werden können. Siehe hierzu: <https://de.wikipedia.org/wiki/Data-Warehouse>. Abgerufen am 21. Oktober 2019.

für die Qualität, die Nutzung und die Bedeutung von Daten in Unternehmen. Technical Data Stewards sind verantwortlich für das Datenmanagement auf Ebene der verwendeten technischen Systeme (Plotkin, 2014, S. 22ff).

Durch die inhärente Bedeutung von Verantwortungsübernahme für die Betreuung von Objekten und deren langfristige Erhaltung, liegt eine Verwendung im informationswissenschaftlichen Kontext, speziell in der Informationsverwaltung und ihrer (archivischen) Bewahrung nah. Informationswissenschaftliche Einrichtungen, insbesondere wissenschaftlichen Bibliotheken, wird eine tragende Rollen in der Betreuung von Daten, die als Forschungsdaten Grundlage oder Produkte von wissenschaftlichen Aktivitäten sind, zuteil (Kindling, Schirmbacher und Simukovic, 2013). Bibliotheken übernehmen nicht nur die Speicherung und Bereitstellung von Forschungsdaten, sondern arbeiten auch bei der Planung des Forschungsdatenmanagements mit Wissenschaftler*innen zusammen (z.B. Tenopir et al., 2014; Yu, 2017). Stewardship in Bezug auf wissenschaftliche Materialien wird in der Literatur in den Kontext von digitalen Artefakten (Dokumente, Daten) gesetzt und als *Digital Stewardship* (National Digital Stewardship Alliance, 2014), *Scientific Stewardship* (Duerr et al., 2004) oder auch *Data Stewardship* (Peng et al., 2015) bezeichnet. Sehr häufig wird dem Begriff in den 2000er Jahren „long-term“ vorangestellt, was eine enge semantische Relation zu Begriffen wie „preservation“ verdeutlicht.

Die National Oceanic and Atmospheric Administration (NOAA) veröffentlichte 2005 einen Plan für ein „*Scientific Data Stewardship Program*“ für Klimadaten (NOAA, 2005). Ziel des Programms war es, Klimadaten langfristig in einer hohen Qualität zu erhalten und für die weitere Forschung zur Verfügung zu stellen. Scientific Stewardship wird im Plan nicht konkret definiert, umfasst jedoch „acquisition and quality control; archival and access for research, business, and government users.“ (NOAA, 2005, S. 26). Das Programm beschäftigt sich inhaltlich mit der Notwendigkeit eines planvollen Umgangs mit Klimadatensätzen, um eine zukünftige Verwendung zu ermöglichen. Qualitätskontrolle, Archivierung und Bereitstellung sind die wesentlichen Elemente bei diesem Verständnis von Stewardship.

Die amerikanische Association of Research Libraries führte 2006 zusammen mit der National Science Foundation (NSF) einen Workshop zu „*Digital data stewardship and preservation*“ durch (Association of Research Libraries, 2006). Im Workshopbericht wird Stewardship als ein Konzept beschrieben, das sowohl Curation als auch Preservation umfasst.

„Stewardship of digital resources involves both preservation and curation. Preservation entails standards-based, active management practices that guide data throughout the research life cycle, as well as ensure the long-term usability of these digital resources. Curation involves ways of organizing, displaying, and repurposing preserved data.“ (Friedlander und Adler, 2006, S. 18)

Stewardship ist in dieser Definition also eine aktive Tätigkeit, die Daten einerseits während des gesamten Forschungskreislaufs betreut und auf eine Langzeitverfügbarkeit vorbereitet und andererseits auf ihre Organisation und Darstellung gerichtet ist, um eine Nachnutzbarkeit in der Zukunft zu garantieren.

Wenn der Begriff in Bezug auf digitale Langzeitarchivierung verwendet wird, liegt seine Bedeutung häufig in diesem Spektrum, wobei der Fokus auf der langfristigen Verfügbarkeit liegt. Teilweise wird Stewardship sogar mit Preservation gleichgesetzt: „*Stewardship: activities that enable use and long-term accessibility of information; often used interchangeably with preservation.*” (Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010). Die National Academy of Sciences (US) et al. definieren in diesem Sinne Stewardship als

„[...] the long-term preservation of data so as to ensure their continued value, sometimes for unanticipated uses. Stewardship goes beyond simply making data accessible. It implies preserving data and metadata so that they can be used by researchers in the same field and in fields other than that of the data's creators. It implies the active curation and preservation of data over extended periods, which generally requires moving data from one storage platform to another. The term “stewardship” embodies a conception of research in which data are both an end product of research and a vital component of the research infrastructure.” (Sciences (US), 2009, S. 27)

Hierbei steht Preservation weiterhin stark im Fokus, jedoch ist mit dem Begriff auch gleichzeitig Curation gemeint. Forschungsdaten sind in dieser Definition sowohl „statische“ Endprodukte von Forschungstätigkeit, die einer Langzeitarchivierung bedürfen, als auch Grundlagen für weitere Forschungszwecke.

Im Unterschied zu Data Stewardship hat sich für die digitale Langzeitarchivierung (LZA) der Begriff *Digital Stewardship* herausgebildet. Die National Digital Stewardship Alliance¹⁶ in den USA ist ein Konsortium von 220 Organisationen, die die Langzeitarchivierung (Preservation) digitaler Information garantieren will und entsprechende Infrastrukturen, Prozesse und Support-Services aufbauen soll. Hierbei geht es auch, aber nicht nur, um (Forschungs-)Daten. Andere Objekte für die LZA sind *electronic records* (wie z.B. Dokumente), Web- und Social Media-Inhalte und digitale Ton- und Bildquellen (National Digital Stewardship Alliance, 2014).

Karasti, Baker und Halkola (2006) beschreiben einen heterogenen Ansatz von Data Stewardship im „Long Term Ecological Research“ (LTER). Sie grenzen den Begriff von ihrer eigenen Definition von Data Curation ab, die ihrer Meinung nach aufgrund eines Fokus auf die technologische Obsoleszenz zu sehr „*technology driven*“ ist und nur „*ingest, archiving und delivery*“ umfasst. Data Stewardship ist im LTER jedoch „*science-driven*“ und wird definiert als

„[...] large conceptual framework, an over-arching process occurring now but attending to the past and taking into account and influencing the future, stretching from data planning to sampling, from data archive to use and reuse – including both data care and information infrastructure work. Such work involves data definitions, data requirements, and quality assurance as well as user feedback, redesign, and data exchange.” (Karasti, Baker und Halkola, 2006, S.

¹⁶Siehe: <https://ndsa.org/about/>. Abgerufen am 21. Oktober 2019.

352)

Diese Definition beinhaltet zwar weiterhin den Bezug zu langfristiger Verfügbarkeit von Daten, nimmt jedoch viel umfassender die aktivere Perspektive von infrastrukturellen Anforderungen und Aspekten der konkreten Organisation von Datenmanagement in den Blick. Strukturen und Prozesse, Policies und Systeme, sowie ein Verständnis von Daten als nicht-statische Objekte sind auch für Henty (2008) die Unterscheidungsmerkmale von Data Stewardship in Abgrenzung zu Digital Curation. Die Definition von Data Stewardship geht über die Kombination von Preservation und Curation hinaus und legt den Fokus weniger auf die Langzeitarchivierung als auf integrierte strukturelle und organisatorische Aufgaben. Baker und Chandler beschreiben in diesem Sinne Data Stewardship als einen holistischen Blick auf das Datenmanagement, der „*data flow and transformation*“ und in diesem Kontext Datenorganisation, Präsentation und Integrität betrachtet (Baker und Chandler, 2008). Hierbei wird die Perspektive eingenommen, dass Daten konstanten Flüssen unterliegen und nie abschließend kuratiert sind. Baker und Yarmey verstehen Data Stewardship als „*a conceptual framework for envisioning the flow of data amongst and between arenas*“ (Baker und Yarmey, 2009, S.14).

Rosenbaum definiert Data Stewardship als

„[...] an approach to the management of data, particularly data, however gathered, that can identify individuals. Data stewardship can be thought of as a collection of data management methods covering acquisition, storage, aggregation, and de-identification, and procedures for data release and use. The concept of a data steward is intended to convey a fiduciary (or trust) relationship with data that turns on a data manager whose loyalty is to the interests of individuals and entities whose data are stored in and managed by the system. Data governance is defined as the process by which stewardship responsibilities are conceptualized and carried out, that is, the policies and approaches that enable stewardship.“ (Rosenbaum, 2010, S. 1444)

Diese Definition bezieht sich vor allem auf Stewardship von medizinischen Daten, die häufig personenbezogene, sensible Daten enthalten. Rosenbaums Definition bezieht neben der Verantwortungsübernahme auch explizit den wichtigen Aspekt des Vertrauens mit ein. Neben der Archivierung, wie bei der Association of Research Libraries und dem geöffneten Blick eines holistischen Datenmanagements wie bei Karasti, Baker und Halkola und Henty, werden hier die zwischenmenschlichen Beziehungen hervorgehoben. Gerade wenn Daten zur Veröffentlichung und (Nach-)Nutzung vorbereitet werden sollen, ist ein vertrauensvoller Umgang mit ihnen und eine gute Beziehung zwischen Datengebern und Data Stewards wichtig (Addison, Moore und Hudson-Vitale, 2015).

Neben Curation und Preservation verschiebt sich der Fokus der Definition von Data Stewardship auch bei Li et al. verstärkt in Richtung Bereitstellung, Publikation und Verfügbarkeit. „[...] [E]fforts to make the data accessible, including the provision of data documents, the linking of data to a digital literature library, the preparation of ready-to-use model data sets and re-analysis data sets, and the provision of a number of Web services“ werden hierbei

ebenfalls unter den Begriff Data Stewardship gefasst (Li et al., 2011, S. 356). Zudem ist hierfür ein extensives Wissen über Anforderungen und Leitlinien (Policies) zu legalen und ethischen Aspekten des Teilens von Daten wichtig (Rolando et al., 2013).

Der National Research Council (NRC) definierte „Scientific Data Stewardship“ 2007 folgendermaßen: „*Scientific data stewardship encompasses all activities that preserve and improve the information content, accessibility, and usability of data and their associated metadata.*“ (National Research Council, 2007, S. 5) Der NRC formulierte in seinen Empfehlungen für die NOAA neun Prinzipien, wobei Principle 6 Stewardship explizit fordert: „*Data and metadata require expert stewardship.*“. Wichtig für nachfolgende Überlegungen zu Scientific Stewardship ist, dass in der Definition des NRC eben nicht nur die Langzeitverfügbarkeit und Kuration der Daten an sich, sondern vor allem auch der Metadaten hervorgehoben wird. Für den NRC ist die Integrität und die zukünftige Integrationsfähigkeit und somit Nachnutzbarkeit für die wissenschaftliche Forschung besonders wichtig. Dies wird sichergestellt durch den Einsatz von Stewards, die die Datensätze kennen und verstehen und an einer Schnittstelle zwischen Datengebernden und Datennutzenden aktiv sind, um die ihnen zugeordneten Datensätze zu pflegen, sowie mit Hilfe von Metadaten die Auffindbarkeit und das Nachnutzungspotential zu garantieren. Stewardship ist hierbei ein iterativer Prozess, der zwischen menschlichen Akteuren, IT-Anforderungen, Speichersystemen, Software, Interfaces und Analysesystemen vermittelt (vgl. ebd., S. 42ff).

Für die Entwicklung eines Reifegradmodells um Scientific Stewardship innerhalb des Departments für National Oceanic and Atmospheric Administration (NOAA) zu messen, haben sich aufbauend auf dieser Definition Peng et al. umfassend mit Funktionen und Implikationen des Scientific Data Stewardships auseinander gesetzt. Sie identifizierten sechs „*non-functional requirements*“ für Scientific Data Stewardship in den Umweltwissenschaften, die durch Policies ein Framework für das Data Stewardship bilden (Peng et al., 2015).

Die in Abbildung 1 dargestellten Implikationen eines „guten“ Data Stewardships umfassen nach Peng et al. die Langzeitverfügbarkeit (Preservability), die Zugänglichkeit und Nutzbarkeit (Accessibility/Usability), die Nachhaltigkeit (Sustainability), die Qualität (Data Quality), die Nachvollziehbarkeit und Reproduzierbarkeit (*Transparency/Traceability/Reproducibility*) und die Integrität (*Information Integrity*) von Forschungsdaten. Die Autoren heben neben den bereits bekannten Funktionen von Data Stewardship (Langzeitverfügbarkeit und langfristige Zugänglichkeit) insbesondere die Zuständigkeit für die Integrität und ein Qualitätsmanagement der Daten hervor, welches erst dazu führen kann, dass Daten reproduzierbar und langfristig nachnutzbar sind. Sie benennen dieses „Scientific Stewardship“ als ein Teilbereich des Stewardships von wissenschaftlichen Ergebnissen (Daten).

„Scientific stewards may provide information or guidance on data quality and characterization of the data to users and may also provide scientific oversight to ensure the accurate scientific representation of values. [...] Therefore, recognizing the role of scientific stewards in caring for scientific data is an important step forward in ensuring data quality and improving usability.“ (Peng et al., 2015, S.233)

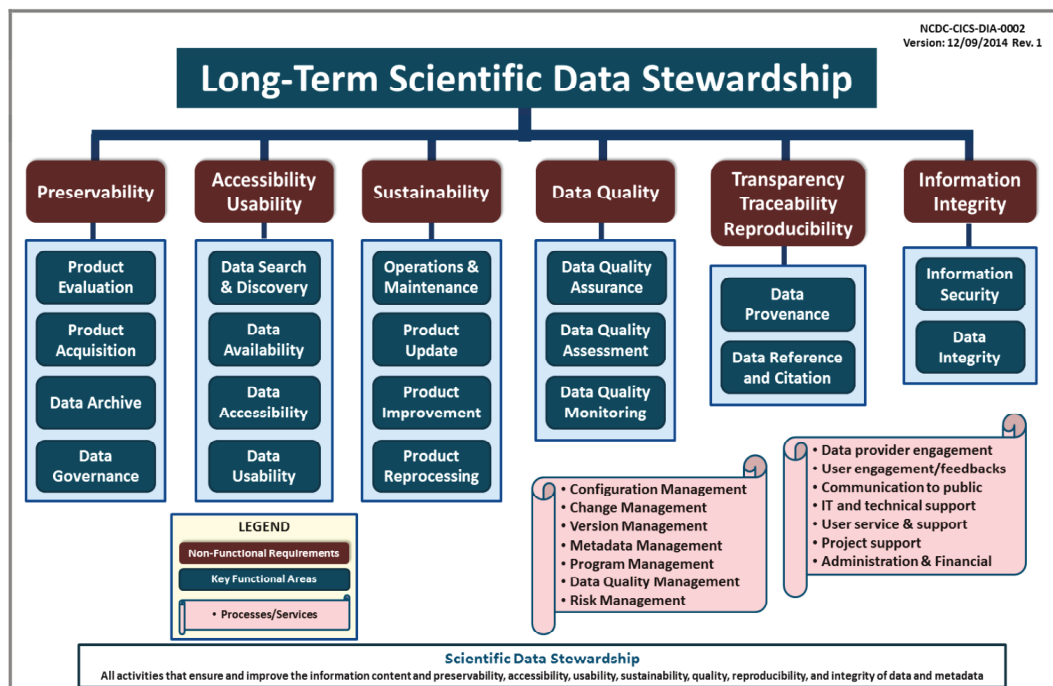


Abbildung 1: Scope des Data Stewardships für Umweltdaten. Aus Peng et al., 2015, S. 236

In der Weiterentwicklung ihres Verständnisses von Stewardship im Datenlebenszyklus identifizierten Peng et al. drei unterschiedliche Ausprägungen des Stewardships, die jeweils als „domain subject matter experts“ in ihrem Tätigkeitsbereich ein hochspezialisiertes Wissen mitbringen (Peng et al., 2016):

- **Data Stewards** sind verantwortlich dafür, dass entsprechende Standards im Datenmanagement, wie Qualitätsstandards oder Policies, eingehalten werden. Zudem leiten sie andere Stewards an und unterstützen bei der Definition von Anforderungen für das Datenmanagement sowie für entsprechende Stakeholder im wissenschaftlichen Team.
- **Scientific Stewards** verfügen über Expertenwissen bezogen auf Inhalte und wissenschaftliche Kontexte eines Datensets. Sie sind verantwortlich für die wissenschaftlich korrekte Repräsentation des Datensets und seiner Metadaten (*scientific integrity*) und deren Qualität. Außerdem unterstützen sie die Entwicklung von Qualitäts- und Nutzbarkeitsanforderungen für andere Stewards und Teammitglieder.
- **Technology Stewards** sind verantwortlich für die technologischen Grundlagen um qualitativ hochwertige und nutzbare Daten und Metadaten zu produzieren. Sie stellen Anleitungen für Software und Systeme zur Verfügung und stellen sicher, dass Anforderungen an Interoperabilität und Integrität auf technischer Ebene der Datensets gegeben sind. Sie unterstützen die Entwicklung von Systemanforderungen an das Datenmanagement zusammen mit anderen Stewards und Teammitgliedern.

Die Autoren verstehen die unterschiedlichen Stewardship-Rollen, wie bereits aus deren Definitionen sichtbar wird, jedoch nicht als voneinander abgegrenzt. Data Stewards können Personen sein, die bereits als Data Manager arbeiten, müssen sich jedoch neben den vorhan-

denen Kenntnissen domänenspezifisches Wissen und technologische Kompetenzen aneignen. Als Scientific Stewards können z.B. die Urheber*innen der Daten agieren, jedoch nur, wenn sie ihre Expertise zusätzlich durch Grundlagen des (langfristigen) Datenmanagements und dessen technologische Anforderungen erweitern und eng mit Personen zusammenarbeiten, die diese Aufgaben übernehmen (z.B. aus der Infrastruktur). Technology Stewards sind meistens Systementwickler*innen u.ä., müssen jedoch, um die Rolle eines Stewards einnehmen zu können, domänenspezifisches Fachwissen und Kenntnisse des (langfristigen) Datenmanagements mitbringen. Diese Definition von Stewardship macht eine spezifische Transdisziplinarität des Konzeptes deutlich. Stewardship geht über (fachspezifische) Tätigkeiten hinaus und setzt die Auseinandersetzung sowohl mit fachlichen und inhaltlichen Aspekten der Forschungsdaten, als auch mit Anforderungen auf technologischer und Management-Ebene (wobei hier vor allem die langfristige Verfügbarkeit gemeint ist) voraus.

Weiterführend entwickelten Peng et al. ein „*Conceptual Enterprise Scientific Data Stewardship Framework* (ESDSF)“. Das ESDSF soll es möglich machen, Data Stewardship Prozesse kontinuierlich weiter zu entwickeln und mit Anforderungen an die erwähnten Tätigkeiten abzugleichen. Stewardship-Aktivitäten, die auf Ebene von einzelnen Datensets angewendet werden, sollen fortlaufend evaluiert werden und es muss ein anhaltendes, transparentes und datenbezogenes Qualitätsmanagement betrieben werden. Das ESDSF beruht auf dem in der Management-Theorie erprobten Plan-Do-Check-Act (PDCA) Zyklus und soll eine daten-zentrische Sicht auf Scientific Data Stewardship vermitteln. Das Framework nimmt dabei zwei Perspektiven ein: Scientific Data-Stewardship als ein systematisches, wissenschaftlich fundiertes Stewardship von Daten („*systematic, scientific way of stewarding data*“) und Scientific-Data Stewardship, als der Umgang mit wissenschaftlichen Daten („*the stewardship of scientific data*“) (Peng et al., 2018, S. 4ff). Die Unterscheidung zwischen Stewardship als eigenständige wissenschaftliche Tätigkeit und dem Stewardship von Ergebnissen wissenschaftlicher Tätigkeit (Daten) ist an dieser Stelle interessant, weil sie auf unterschiedliche Verantwortlichkeits- und Rollenverständnisse im Umgang mit wissenschaftlichen Daten hinweist und Stewardship sowohl als Teil der wissenschaftlichen Arbeit, als auch als eigene Aufgabenbeschreibung ansieht. Für die Autoren führt Stewardship - als Kombination aus Data, Scientific und Technology - zu qualitativ hochwertigen, nachhaltigen und nutzbaren Datenprodukten, die zudem den höchsten (fach-)wissenschaftlichen Standards und Anforderungen entsprechen.

Wilkinson et al. publizierten die FAIR-Prinzipien im direkten Zusammenhang mit Data Stewardship. In ihrem weitreichend zitierten Artikel „The FAIR Guiding Principles for scientific data management and stewardship“ wird Stewardship folgendermaßen beschrieben:

„Beyond proper collection, annotation, and archival, data stewardship includes the notion of ‘long-term care’ of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies.“ (Wilkinson et al., 2016, S. 1)

In dieser Definition nähert sich Stewardship eher wieder älteren Definitionen an. Das Sammeln, Beschreiben und Archivieren von Daten sind Aktivitäten, die sich unter Preservation kategorisieren lassen. Auch die Bezeichnung „*long-term care*“ erinnert an die LZA. Der Unterschied zu früheren Definition aus der Sichtweise der LZA ist bei Wilkinson et al. die Beschreibung des Ziels von Data Stewardship. Langzeitverfügbarkeit von Daten ist nicht gleichzusetzen mit der in dieser Definition beschriebenen klaren Ausrichtung auf Auffindbarkeit, Evaluierbarkeit und Nachnutzbarkeit der Daten. Daten sollen demnach nicht nur verfügbar sein, sondern durch ein long-term Stewardship langfristig interoperabel und nachnutzbar vorliegen (European Commission, 2018b).

Im Zuge der Entwicklung der EOSC wird der Begriff Data Stewardship mit dem Bezug zu FAIR Data sehr häufig verwendet. Data Stewardship wird dabei als die Verantwortungsübernahme für eigene und fremde Daten über den Forschungsprozess und darüber hinaus betrachtet (European Commission, 2016). Eines der Ziele der EOSC ist „*FAIR data management and tools to ensure data stewardship across borders and disciplines*“ zu entwickeln (European Commission, 2018c, S. 16). Besonders bei der „FAIRifizierung“ von Daten werden gut ausgebildete Data Stewards benötigt, die als „Support Staff“ aus Forschungscommunities, wissenschaftlichen Informationsinfrastrukturen (z.B. Bibliotheken) oder aus Service-Infrastrukturen (z.B. Repositorien) stammen können. Stewardship ist dabei „*a set of skills to ensure data are properly managed, shared and preserved throughout the research lifecycle and in subsequent storage*“ (Collins et al., 2018, S. 13).

Durch die Formulierung der FAIR Prinzipien und damit einem sehr umfassenden Erwartungshorizont an nachnutzbaren, interoperablen Daten, wird auch der Begriff Stewardship zunehmend als eine kollektive Tätigkeit, die eine enge Kollaboration zwischen Forschenden, Informationsinfrastruktur und Fachcommunities betrifft, verstanden. Whyte et al. fassen diese Sicht folgendermaßen zusammen:

„As the FAIR principles articulate, stewardship implies long-term and sustainable care across multiple lifecycles. This makes stewardship a collective endeavour, involving at least the individual researcher, colleagues in the study ('the team'), their host organisation (or and others providing services) and the research domain(s) or communities that care about the data.“ (Whyte et al., 2018, S. 29)

Die Autoren definieren Stewardship dementsprechend als „*the formalisation of roles and responsibilities and their application to ensure that research objects are managed for long-term reuse, and in accordance with FAIR data principles.*“ (ebd., S. 11)

Am Dutch Techcenter for Lifesciences (DTL) sind sogenannte „Data Stewards“ angestellt, die vor allem im Bereich der Lebenswissenschaften arbeiten. DTL arbeitet in enger Zusammenarbeit mit ELEXIR, einer europäischen Organisation für das Management von lebenswissenschaftlichen Forschungsdaten¹⁷, an Trainings und Integrationsmöglichkeiten von Data Stewardship in den Forschungsprozess und die FAIRifizierung der Forschungsergebnisse.

¹⁷Siehe: <https://elixir-europe.org/>. Abgerufen am 21. Oktober 2019.

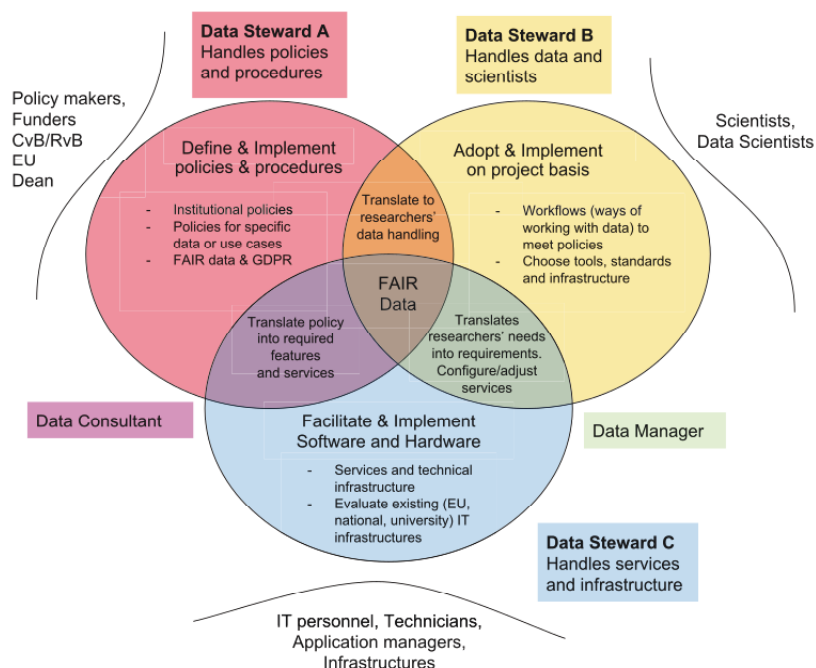


Abbildung 2: Darstellung der Stewardship Dimensionen des DTL. Aus Scholtens et al., 2019b

Beispielsweise wurde der Data Stewardship Wizard¹⁸ für die vereinfachte Erstellung von DMPs für FAIR-konforme Daten entwickelt. Data Stewards am DTL sind „[r]esponsible for planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data. Data stewards help researchers to maintain their data. Data stewards maintain data even if researchers left the institute.“ (Staiger, 2019) Im einem Projekt wird zusammen mit anderen niederländischen Forschungseinrichtungen eine Matrix für die konkreten Funktionen von Data Stewards in den Lebenswissenschaften entwickelt (Scholtens et al., 2019a). Ein Ergebnis ist, dass Data Stewards im Spannungsverhältnis von wissenschaftspolitischen Anforderungen (Policies, Anforderungen der Forschungsförderer usw.), Forschenden und IT drei unterschiedliche Dimensionen abbilden können, die sich jeweils an ihren Grenzen überschneiden und an ihrer gemeinsamen Schnittstelle FAIR-konforme Daten als Ziel haben (siehe Abbildung 2, vgl. Scholtens et al., 2019b):

- **Data Steward A** arbeitet auf der Ebene von Policies und darauf abgestimmten Prozessen. An der Schnittstelle zu Technologien werden diese Anforderungen in IT-Services übersetzt, indem zum Beispiel spezifische Software entwickelt oder bereit gestellt wird. An der Schnittstelle zu Nutzenden und Datenproduzent*innen werden diese Anforderungen in den Umgang mit Daten übersetzt. Data Stewards vom Typ A werden von Akteuren auf einer administrativen Ebene beeinflusst.
- **Data Steward B** arbeitet auf Ebene des aktiven Umgangs mit Daten, Tools und Workflows. An der Schnittstelle zu Policies wird dieser Umgang auf entsprechende Anforderungen abgestimmt. An der Schnittstelle zu Technologien werden Bedürfnisse

¹⁸Siehe: <https://ds-wizard.org/>. Abgerufen am 21. Oktober 2019.

von Forschenden für IT-Services übersetzt. Data Stewards vom Typ B werden von Akteuren aus den Forschungscommunities beeinflusst.

- **Data Steward C** arbeitet auf der Ebene von technologischen Infrastrukturen. An der Schnittstelle zu Policies werden diese Services entsprechend angepasst. An der Schnittstelle zum aktiven Umgang mit Daten werden diese Services entsprechend den Bedürfnissen der Forschenden angepasst. Data Stewards vom Typ C werden von Akteuren aus den technischen Infrastrukturen beeinflusst.

Interessant ist bei dieser Darstellung der Vergleich zu der von Peng et al. entwickelten Kategorisierung von Stewardship. Typ B wäre hierbei mit Scientific Stewardship und Typ C mit Technology Stewardship gleichzusetzen. Typ A kann in etwa mit Data Stewards gleichgesetzt werden, obwohl bei Scholtens et al. administrative Anforderungen mehr im Vordergrund stehen als bei Peng et al.

Kontext	Beispiele	Ziel
Digitale Langzeitarchivierung, Data Curation	NOAA, 2005; Friedlander und Adler, 2006; Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010	Langzeitverfügbarkeit
Life-Cycle, Daten als lebendige Objekte	Karasti, Baker und Halkola, 2006; Henty, 2008; Baker und Chandler, 2008	Aktives Datenmanagement
Vertrauen, Data Security	Rosenbaum, 2010; Addison, Moore und Hudson-Vitale, 2015	Aktives Datenmanagement
Bereitstellung, Data Sharing	Li et al., 2011; Rolando et al., 2013	Aktives Datenmanagement
Qualitätskontrolle, Usability	National Research Council, 2007; Peng et al., 2015; Peng et al., 2016; Peng, 2018	Aktives Datenmanagement
Interoperabilität, Nachnutzung	Wilkinson et al., 2016; Whyte et al., 2017; Whyte et al., 2018; European Commission, 2016; Hodson et al., 2018	FAIR Data

Tabelle 1: Begriffsentwicklung Stewardship. Eigene Darstellung.

Zusammenfassend lässt sich in der Literatur somit eine Entwicklung der Begriffsdefinition von Data Stewardship ausmachen. In Tabelle 1 ist dargestellt, dass sich die Definition von (Data) Stewardship Langzeitverfügbarkeit von (statischen) „Datenprodukten“ im Kontext der LZA, in den letzten Jahren eher eine dem Kontext „aktives Datenmanagement“ zugeordnete Bedeutung entwickelt hat. Mit dem Aufkommen der FAIR Data Principles wurden den beiden

vorherigen Bedeutungen noch das Ziel der Interoperabilität und Nachnutzbarkeit von Daten hinzufügt. Zudem zeichnen sich unterschiedliche Ausprägungen von Data Stewardship ab, die Technologien, wissenschaftspolitische Anforderungen und disziplinspezifische Besonderheiten in den Blick nehmen.

Der Begriff Data Stewardship ist nicht neu, erfährt jedoch im engen Bezug zu den FAIR Data Prinzipien einen fast inflationären Gebrauch. Die Auswertung der begrifflichen Entwicklung zeigt, dass sich jedoch in den letzten Jahren Kontexte und Ziele stark erweitert haben. Data Stewardship verortet sich hiernach im Sinne des aktiven Datenmanagement an Interaktionspunkten zwischen unterschiedlichen (Fach-)Domänen, die am Datenmanagement beteiligt sind. Diese Verortung betont den transdisziplinären Charakter der Tätigkeit sowie eine Schwellenfunktion zwischen verschiedenen Akteursgruppen. Ziele des Data Stewardships sind nicht mehr nur die langfristige Bewahrung von Forschungsergebnissen sondern vielmehr die Bereitstellung von wissenschaftlichen Daten, die sowohl für Menschen als auch für Maschinen verständlich und nachnutzbar vorliegen.

4.2 Begriffsspektrum

Allgemein muss bei der Betrachtung von Stewardship unterschieden werden zwischen

1. Stewardship als Konzept,
2. Stewardship als praktische Tätigkeit und
3. (Data) Steward als Rollen- oder Berufsbezeichnung.

Wenn Stewardship als **Konzept** beschrieben wird, geschieht dies meistens in Verbindung mit oder in Abgrenzung zu Data Management und Data Curation. Dabei fällt auf, dass alle drei Konzepte und die Unterscheidungen zwischen ihnen in der Literatur nicht klar definiert sind. Palmer et al. (2013) sehen Data Curation beispielsweise in Abgrenzung zu Data Stewardship als eine Aufgabe, bei der die Aufbereitung der Daten für zukünftige Forschungsinteressen im Vordergrund steht. In ihrer Definition von Data Curation ist die Angleichung von Tätigkeiten an den Forschungsprozess das Merkmal, das Data Curation von Data Stewardship abgrenzt. Data Stewardship ist hingegen „*a function of 'managing data' that implies a less active, fixed maintenance of data over time*“ (Palmer et al., 2013, S. 7).

Gegensätzliche Meinungen sehen Data Stewardship im Vergleich zu Data Curation als breiter gefasstes Konzept. Aus dieser Sicht befasst sich Curation eher mit einer kurzweiligen Sicht auf Daten, während Data Stewardship eine ganzheitliche und langfristige Perspektive auf Datensätze als dynamische Elemente beinhaltet (Karasti, Baker und Halkola, 2006). Aus der Sicht von Service-Anbietern ist Curation eher ein lokal bezogenes Konzept, beispielsweise im Sinne der Aufbereitung von Daten für ein lokales Repositorium, während Stewardship vernetzte Infrastrukturen und heterogene Datenspeicherungen und -verwertungen betrachtet (Baker und Yarmey, 2009).

Aus konzeptioneller Sicht hat sich Data Stewardship in den letzten Jahren entsprechend

der Begriffsdefinitionen aus Tabelle 1 ausgebildet. Der Kern von Data Stewardship ist im heutigen Gebrauch „*[the] notion of ‘long-term care’ of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data*“ (Wilkinson et al., 2016, S. 1), wobei Stewardship kein Ziel an sich ist, sondern den Prozess und die Grundlage für die Bereitstellung von FAIR-konformen Daten beschreibt.

Konkrete **Praktiken**, die Stewardship als ein Menge von zielgerichteten Aktivitäten beschreiben, sind in der Literatur dagegen, gerade in Abgrenzung zu Management und Curation Praktiken, häufig nicht genau definiert. Duerr et al. beschreiben Tätigkeiten im Bereich von Data Collection, Sharing, Publishing, Preserving, Metadata Management, Formats, Preservation, Access und Data Security and Integrity (Duerr et al., 2004). Peng et al. geben folgende Aufgaben an: „*manage, curate, and serve data to meet user demand, curating metadata, ensuring data fixation, und supporting data access*“ (Peng et al., 2018, S. 1). Besonders in der Entwicklung der EOSC und im Hinblick auf FAIR-konforme Daten, werden Praktiken des Data Stewardship momentan weiter konkretisiert. Dies geschieht auch vor dem Hintergrund benötigter Expert*innen und Kompetenzen für die Umsetzung der EOSC. Data Stewardship beinhaltet Aktivitäten wie Datenbereinigung während des Forschungsprozesses, Strukturierung von Datensätzen, die Erstellung von Metadaten und die adäquate Speicherung von Daten. In Hinblick auf das Teilen und die Nachnutzung von Daten gehören Tätigkeiten wie die Dokumentation von Daten, die Sicherstellung der Langzeitverfügbarkeit und die Bereitstellung von Datensätzen zu den Verantwortlichkeiten im Data Stewardship. Außerdem fallen Aufgaben wie die Entwicklung von Standards, Best Practices und Frameworks für die Interoperabilität von Daten an (Collins et al., 2018).

Konkrete Tätigkeiten wurden auch im Projekt EOSCpilot untersucht, in dem Kompetenzen und Fähigkeiten von Data Stewards erarbeitet wurden. Diese sind in Kapitel 6 genauer dargestellt. Data Stewardship wird hier als eine Schwellenfunktion zwischen (fach-)wissenschaftlichen Tätigkeiten, Data Science und Datenanalyse, der Entwicklung von Daten-services und Datenmanagement und -kuration beschrieben. Die verzögerte Entwicklung eines solchen Frameworks zeigt jedoch auch, dass Stewardship bislang eher auf konzeptioneller Ebene betrachtet wurde und mit einer Reihe von implizit vorhandenen, aber nie explizit gemachten Tätigkeiten verknüpft war. Im Zuge der Entwicklung von neuen Rollen-, Aufgaben- und sogar Berufsbezeichnungen wird die Notwendigkeit von Tätigkeitsbeschreibungen und die Einbettung von Praktiken in Prozesse erst evident.

Ein im Juni 2019 erschienener Bericht zu Data Stewardship Entwicklungen in mehr als 30 niederländischen Forschungseinrichtungen durch den die niederländische nationale Forschungsdaten-Koordinationsstelle (LCRDM) zeigt das Ergebnis einer Auswertung von Stellenausschreibungen aus dem erweiterten Bereich des Datenmanagements in den Niederlanden und klassifiziert drei Bereiche für die Aufgaben des Data Stewardship in Einrichtungen (Verheul et al., 2019, siehe Abbildung 3):

- *Embedded and operational*
- *Generic and advice*
- *Policy, strategy and coordination*

Mit *embedded and operational* sind dabei Aufgabenbereiche, die sich mit dem konkreten und direkten Umgang mit Daten befassen, gemeint. Dies kann beispielsweise statistische Analysen, den Aufbau und die Betreuung von Datenbanken, die Extrahierung und Prozessierung von Daten, Workflow- und Prozessentwicklungen für technologische Infrastrukturen oder die Nachnutzung und/oder Entwicklung von Algorithmen umfassen. Diese Aufgabenbereiche zeichnen sich durch hohe Kompetenzanforderungen im Bereich der angewandten Datenwissenschaft (Data Science) aus.



Abbildung 3: Aufgabenbereiche im Data Stewardship. Aus Verheul et al., 2019, S. 8.

Aufgaben im Bereich der *generic and advice*-Kategorie umfassen eher Tätigkeiten in der Beratung und Unterstützung im Datenmanagement von Angehörigen der Einrichtungen. Beispielsweise fallen hierunter die Beratung von Forschenden und Doktoranden zu Datenmanagement-Standards und Qualitätsaspekten, die Vernetzung mit Forschungsteams, die Zusammenarbeit mit potentiellen Multiplikatoren in den jeweiligen Fakultäten, die Entwicklung und Durchführung von Workshops oder Kursen zum Datenmanagement und Networking-Tätigkeiten im größeren Bereich der Open Science.

Eine Betonung von *policy, strategy and coordination*-Tätigkeiten wurde weniger häufig in den untersuchten Stellenausschreibungen festgestellt, wurde jedoch als dritte Kategorie aufgenommen, da es sich um Tätigkeiten handelt, die sich den ersten beiden Kategorien nicht gut zuordnen lassen. Aufgaben umfassen hier strategische und taktische Entwicklungen von Services im Bereich des Datenmanagements, die Entwicklung von Leitlinien zum Umgang mit Forschungsdaten (Data Policies) an den Einrichtungen, das Übertragen von übergeordneten Policies auf die jeweiligen Einrichtungen und die Funktion als Mediator*in zwischen fachwissenschaftlichem und administrativem Personal an den Einrichtungen.

Die Untersuchung bezog sich auf eine eher geringe Menge von 22 Stellenausschreibungen, die zwischen Juni 2018 und Februar 2019 in den Niederlanden veröffentlicht wurden. Festzuhalten ist auch, dass keine der Ausschreibungen nur eine Kategorie von Data Stewardship abdeckte, sondern auch Aufgaben aus anderen Bereichen berücksichtigten. Trotz einiger Schwächen, wie z.B. der geringen Datenbasis, zeigt die Studie, dass Data Stewardship unterschiedliche Aufgabenbereiche umfasst, die sich jedoch wie von Scholtens et al. (2019a) gezeigt, in drei

Richtungen ausprägen kann. Eine Umfrage unter für das Datenmanagement verantwortlichen Personen folgte der Auswertung der Stellenausschreibungen und zeigte, dass Tätigkeiten aus den Bereichen generisches Datenmanagement und im Bezug auf strategische Planungen bereits größtenteils so erfüllt werden. Bedarf besteht vor allem noch in der Standardisierung und Integration von Daten (Verheul et al., 2019, S. 14). Auffällig an den Ergebnissen der Umfrage ist, dass Data Curation relativ häufig mit „*not available*“ angegeben wird, Data Management jedoch in fast allen befragten Einrichtungen ausgeführt wird. Dies könnte unter Umständen an unterschiedlichen Interpretationen des Begriffs Curation liegen, da in der Umfrage nicht erklärt wird, was genau mit den Begriffen gemeint ist und diese, wie bereits oben erwähnt, häufig nicht klar voneinander abgrenzbar sind.

Aus den in der Umfrage vorgegebenen Aufgaben für Data Stewards wurden folgende sechs Aufgaben am häufigsten ausgewählt (vgl. ebd., S. 15):

1. Die Unterstützung von Forschenden
2. Eine Konsultationsfunktion zum Thema Datenmanagement
3. Das Verbreiten der FAIR Data Prinzipien
4. Die Nutzung von vorhandenem Fachwissen und Infrastrukturen
5. Die Entwicklung von und/oder das Durchführen von Trainings zum Thema Datenmanagement
6. Die Kollaboration mit anderen Organisationen und Fakultäten während des Forschungsvorhabens

Generell ist die Umfrage eher vage gehalten. So ist nicht ersichtlich, auf welcher Grundlage die Auswahlmöglichkeiten für die Frage nach Aufgaben im Data Stewardship erstellt wurden. Angesichts der Rolle von Data Stewards im Allgemeinen ist es sicherlich nicht verwunderlich, dass die Unterstützung von Forschenden an erster Stelle steht. Auffällig ist jedoch, dass für die Aufgaben „*Is accountable for data*“ und „*Is responsible for data*“ eine eher geringe Zustimmung besteht. Dies zeugt von einer Funktion der Data Stewards als Vermittler und Berater, anders als ursprüngliche Definitionen von Data Stewardship (z.B. Rosenbaum), die die aktive Verantwortungsübernahme für die Daten anderer beschreiben.

Eine Sicht auf Data Stewardship in Bezug auf eine solche **Rollenverteilung** ist bis dato kaum geschehen. In der Literatur wurde Stewardship bislang als Konzept oder Prozess beschrieben, an dem mehrere Akteure beteiligt sind und dem keine konkrete Berufsbezeichnung zuzuordnen ist. Dies liegt daran, dass Stewardship als dynamisches Konzept unterschiedlichen Akteuren zugerechnet wird. Sowohl Forschende selbst, als auch Anbieter*innen von Forschungsinfrastrukturen können demzufolge Stewardship betreiben. Anders als in der Wirtschaft, wo Data Stewards als Teil des Data Governance Teams schon länger eine genaue Berufsbezeichnung haben (Plotkin, 2014), wurden erst in den letzten Jahren vor allen in den Niederlanden, aber auch in Belgien und Österreich Stellenausschreibungen für Data Stewards veröffentlicht. Data Stewards werden meist durch die Bibliothek koordiniert, sind jedoch angesiedelt an der Grenze zwischen Infrastruktur und Fachwissenschaften. Sie fungieren als „*first point of reference for all data related questions*“ (Shalini Kurapati, 2019). Data Stewards haben eine fachwissenschaftliche Ausbildung und können so zwischen den jeweiligen Anforderungen von Forschenden und Infrastrukturen vermitteln. Da sie nicht per se in der

Bibliothek, sondern vielmehr in den jeweiligen Fachbereichen arbeiten, kann diese Rolle auch als *embedded* Data Stewardship bezeichnet werden (Neuroth et al., 2019). Hierbei ist die Rolle bzw. organisatorische Verortung der Data Stewards in der Einrichtung gemeint und nicht die reine Tätigkeitsbeschreibung wie sie mit „embedded and operational“ in Verheul et al. (2019) beschrieben ist.

Diese Ansiedlung kann in unterschiedlichen Formen geschehen. Data Stewardship kann an verschiedenen Einrichtungen sehr divers organisiert sein. Das Data Stewardship Team kann beispielsweise aus der Kollaboration von Bibliothek, Rechenzentrum und Datenschutzbeauftragten entstehen und durch eine zentrale Stelle koordiniert werden. Die einzelnen Fakultäten entsenden eine*n Data Steward, manchmal werden Stewardship Aufgaben jedoch auch durch entstandene neue Rollen, wie Research Software Engineers oder andere Felder in der IT-Governance erfüllt (Verheul et al., 2019, S. 23f). In einer anderen Einrichtung werden Forschende durch die Bibliothek, ein Forschungsdatenzentrum und entsprechenden Koordinatoren betreut. An der TU Delft wird außerdem zwischen Data Stewards und sogenannten Data Champions unterschieden. Data Stewards sind Festangestellte, die für die einzelnen Fakultäten verantwortlich sind, das Datenmanagement unterstützen, Strategien entwickeln und Trainings anbieten. Data Champions werden von den Data Stewards betreut und arbeiten nur lokal auf Ebene der Forschungsteams. Sie fungieren damit als Multiplikatoren für die Data Stewards (Teperek und Dunning, 2019). An der TU Delft kommen Data Stewards aus unterschiedlichen Disziplinen und sind diesen Fachbereichen auch zugeordnet. Sie beraten Forschende somit, vor einem fachwissenschaftlichen Hintergrund zu operativen Anforderungen und Planung des Datenmanagement, zu Anforderungen von Seiten der Forschungsförderer, Fachzeitschriften und Workflows. Sie führen außerdem Trainings und Schulungen durch, evaluieren bestehende Praktiken und sind beteiligt an der Erstellung von (fachspezifischen) Policies (Teperek et al., 2018).

Die in der Studie untersuchten Stellenausschreibungen verwendeten für den Bereich des „embedded and operational“ Stewards neben Berufsbezeichnungen wie „(Research) Data Manager“ auch „Data Scientist“ und ein Mal die Bezeichnung „Bioinformatician“. Für den Bereich *policy, strategy and coordination* umfassten Berufsbezeichnungen dementsprechende Koordinierungsfunktionen, wie z.B. „Data Stewardship Coordinator“, „Research Data Management Coordinator“ oder „Senior Research Data Manager“. Für „generic and advisory“ wurden Stellenbezeichnungen wie „Data Steward“ oder „Research Data Management Specialist“ identifiziert (Verheul et al., 2019). Damit wird deutlich, dass die Rolle von Data Stewards in Forschungseinrichtungen bereits zumindest teilweise übernommen werden konnte, z.B. von entsprechend ausgebildeten Mitarbeiter*innen in den Bibliotheken oder Rechenzentren, ohne dass diese den Titel Data Steward tragen. Außerdem wird deutlich, dass die Rolle des Data Stewards kaum umfassend von nur einer Person ausgefüllt werden kann. Idealerweise arbeiten Data Stewards in den unterschiedlichen Rollenausprägungen als Team zusammen.

4.3 Verwendung des Begriffs in dieser Arbeit

In dieser Arbeit wird der Begriff Data Stewardship vor dem Hintergrund seiner Begriffsentwicklung als koordinierende und unterstützende Tätigkeit in einer Schwellenfunktion zwischen Forschenden, Infrastruktureinrichtungen und Administration einer Forschungseinrichtung betrachtet. Dabei werden folgende Aspekte berücksichtigt:

1. Das Stewardship von Forschungsdaten umfasst als **Konzept** den gesamten Forschungskreislauf und ist sowohl auf die langfristige Vorhaltung der anfallenden Forschungsdaten als auch auf ihre Bereitstellung, Interoperabilität und Nachnutzbarkeit ausgerichtet. Data Stewardship ist Teil des gesamten Forschungsprozesses und umfasst in dieser Weise auch die Aktivitäten Datenmanagement und Data Curation/Preservation.
2. Data Stewardship ist als Menge an **Tätigkeiten** sowohl Forschenden selbst als auch den forschungsunterstützenden Stellen zuzuordnen. Anders als Data Curation oder Data Preservation beginnt Stewardship bereits bei der Erstellung von Daten und sollte von Beginn des Forschungsvorhabens eingeplant werden. Dabei ist das Schaffen von Berufsgruppen wie Data Stewards eine Entwicklung, in deren Rahmen die Aufgaben professionalisiert und Verantwortlichkeiten besser verteilt werden können.
3. Data Stewards als konkrete **Akteure** sind nicht, wie in ursprünglichen Definitionen von Stewardship, singular verantwortlich für die anfallenden Daten. Sie verfügen über notwendige Fachexpertise und Kompetenzen im Datenmanagement, um fachspezifische Beratungsfunktionen übernehmen zu können, wirken jedoch weniger operativ auf den Umgang mit Daten ein. Vielmehr agieren sie vermittelnd und unterstützend für alle Beteiligten Stewardship-Tätigkeiten.

Unterschieden werden muss zwischen Data Stewardship als umfassende und zusätzliche Aufgabe während der Forschungstätigkeiten *durch Forschende* und Data Stewardship, welches einen Aufgabenbereich von Data Stewards beschreibt. Der Grund dafür ist, dass genaue Tätigkeiten von Data Stewards noch nicht abschließend geklärt sind und Kompetenzanforderungen, wie z.B. im EOSCpilot (Whyte et al., 2018), tatsächlich eher Tätigkeitsbeschreibungen für alle am Datenmanagement beteiligten Akteure darstellen. Im Folgenden wird daher davon ausgegangen, dass Data Stewardship-Tätigkeiten einer Vielzahl von Akteuren zuzuordnen sind und nicht als alleiniger Tätigkeitsrahmen für Data Stewards angesehen werden können. Data Stewards wirken eher als Multiplikator*innen des Data Stewardships und als Integrator*innen des Konzeptes in die ihnen zugeordneten Einrichtungen. Sie beraten und unterstützen alle anderen Akteure im Datenmanagement um eventuell auftretende Probleme zu vermeiden und kennen Spezialist*innen, auf die für spezifische Fragestellungen verwiesen werden kann. Folglich passt die Allegorie für Data Stewardship der TU Delft hier gut: Data Stewards sind Allgemeinärzte („*general practitioners*“) für alle Tätigkeiten des Data Stewardships (Shalini Kurapati, 2019).

5 Boundary Work

5.1 Arten von Boundary Work

Das Begriffsspektrum von Data Stewardship und das Verständnis von Data Stewards als beratende und vermittelnde Akteure deutet bereits darauf hin, dass Stewardship unterschiedliche Akteure involviert und Data Stewards an den Schwellenbereichen zwischen Forschung, Infrastruktur und Forschungsadministration, die als unterschiedliche Soziale Welten angesehen werden können, arbeiten. Wirken an den Grenzen zwischen sozialen Welten wird in der Literatur auch als Boundary-Work (Grenzarbeit/Schwellentätigkeit) beschrieben.

Der Begriff Boundary-Work kann auf unterschiedliche Weise verstanden werden. Aktivitäten, die zur Festigung von Grenzen stattfinden, können als Boundary-Work bezeichnet werden. Diese Abgrenzungsmechanismen werden in der Wissenschaftskommunikation beispielsweise zwischen den Welten Wissenschaft und „Nicht-Wissenschaft“ bzw. Öffentlichkeit beschrieben. Bei dieser Form der Grenzarbeit geht es in erster Linie um das Herstellen von Autoritäten und Machtgefügen, indem sich Wissenschaft als „Expertentum“ klar von der Öffentlichkeit abgrenzt und das Festigen von welt-eigenen Identitäten und Legitimationsstrategien gefördert wird (Gieryn, 1983). Obwohl auch diese Form der Grenzarbeit für die Herstellung funktionierender Sozialer Welten und die Etablierung sozialer Identitäten wichtig ist (Wikström, 2008), wird Boundary-Work im folgenden als „Arbeit an Grenzen“, oder Schwellenarbeit, mit einem Fokus auf Grenzüberbrückungen verstanden. Es kann hier unterschieden werden zwischen „*boundary spanning*“, „*boundary buffering*“ und „*boundary reinforcing*“ (Faraj und Yan, 2009). Während die beiden letztgenannten Konzepte eher auch Abgrenzungsmechanismen in den Blick nehmen, sind Boundary-Spanning- bzw. Boundary-Crossing-Aktivitäten in der grenzübergreifenden Zusammenarbeit von Bedeutung. Boundary Objects können dabei als Werkzeuge für die Grenzarbeit eingesetzt werden.

Boundary-Spanning-Aktivitäten können in drei Kategorien eingeteilt werden: *trading*, *sharing* und *knowing*. *Trading* geschieht in sogenannten „*Trading Zones*“, in denen Informationen aus den unterschiedlichen Welten sichtbar gemacht und gespeichert werden, beispielsweise in Datenbanken (*displaying*). Bei dieser Kollaboration dienen Objekte wie Kalender, E-Mails oder Agenden als BOs. In den BOs wird Wissen repräsentiert, gespeichert und mit ihrer Hilfe abgerufen (*representing*). Über das Teilen der BOs werden Grenzen überwunden und neue Wissensartefakte geschaffen (*assembling*) (Kellogg, Orlikowski und Yates, 2006). *Sharing* bezieht sich auf ein von Carlile (2004) beschriebenes Modell von Transfer, Translation und Transformation. Grenzüberschreitungen werden durch die Entwicklung von gemeinsamen Lexikalitäten, Bedeutungen und Agenden durchgeführt. BOs dienen dabei dazu, soziale Relationen zu verhandeln, Diskurse zu führen und Einstellungen zu verändern (Carlile, 2004, vgl. dazu Kapitel 5.2). Eine andere Form des *sharing* ist das Vermitteln zwischen unterschiedlichen Welten (*brokering*). Die Person, die als „Broker“ identifiziert wird, schafft eine Verbindung zwischen den beteiligten Welten, stellt eine Kombinationen der jeweiligen Sichten im Bezug zu einem BO als Referenzobjekt her, und vermittelt damit Informationen und Wissen zwischen den Welten (Pawlowski und Robey, 2004). Wenn Zusammenarbeit vor dem Hintergrund kollektiver Aktivität innerhalb sozialer Welten betrachtet

wird, können Grenzüberschreitungen zudem als Wissenspraktiken (*knowing*) definiert werden. Wissenspraktiken beziehen dabei auch Lernprozesse mit ein, die bei der grenzübergreifenden Zusammenarbeit entstehen. BOs können dabei implizites Wissen explizit machen und Arten des Wissenstransfers und der Wissensarbeit innerhalb verschiedener Welten offen legen (Bechky, 2003).

Personen, die verbindend an Grenzen arbeiten (sog. *Boundary Spanner*), haben ähnlich vielfältige Qualitäten wie Grenzen selbst. Im Zuge ihrer Brückenfunktion sind sie weder einer Welt komplett zugehörig noch explizit unabhängig. Sie fungieren jedoch nicht nur als Brücke zwischen den einzelnen Welten, sondern repräsentieren auch gleichzeitig ihre Grenzen. Grenzarbeitende sind Übersetzer*innen, Mediator*innen und Verhandelnde zwischen den einzelnen Welten und verbinden diese durch tauschen, teilen und hervorbringen von Wissen: „[...] *They enact the boundary by addressing and articulating meanings and perspectives of various intersecting worlds. At the same time, these people [...] move beyond the boundary in that they have unspecified quality of their own (neither-nor).*“ (Akkerman und Bakker, 2011, S. 11).

Für die an den Grenzen agierenden Personen bedeutet dies die Notwendigkeit einer inneren Reflexion der eigenen Identität im Sinne von reflexiven Abgrenzungsmechanismen („Wer bin ich im Gegensatz zu den anderen?“) sowie eine stark nach außen getragene Kommunikationsbereitschaft mit allen Akteuren (Wikström, 2008). Boundary Spanner müssen komplexe, interagierende Systeme (Interessen, Motivationen, Normen, Personen, Organisationen, ect.) verstehen und einschätzen können. Sie verhandeln zwischen unterschiedlichen Autoritäten und Identitäten mit dem Ziel, Konflikte aufzulösen und Kompromisse zu finden. Dabei besteht ihre Aufgabe auch darin, Informationsflüsse in alle Richtungen zu filtern und zu lenken, um Grenzen in der Kommunikation überwinden zu können. Die Grundlage für ein gelungenes Boundary-Spanning und somit das Schaffen der Möglichkeit von Kollaboration ist hierbei der Aufbau eines Vertrauensverhältnisses mit allen involvierten Parteien (Harris und Lyon, 2013).

Boundary Spanner können dabei unterschiedliche Rollen einnehmen, die als Netzwerkende, Entrepreneur*in, Interpretator*in/Kommunikator*in und Koordinator*in beschrieben werden. Als Netzwerkende verhandeln sie Beziehungen zwischen unterschiedlichen Akteuren. Dabei sind insbesondere interpersonelle, kommunikative, diplomatische und strategische Kompetenzen gefragt. Als Entrepreneur*in sollten Boundary Spanner in der Lage sein, mit Hilfe von Kreativität und analytischem Denken Lösungen und Handlungsstrategien für komplexe Probleme zu entwickeln. Besondere Gewichtung liegt für Boundary Spanner auf den Rolle als Kommunikatoren und Interpreatoren - die auch als eine Art „Mediatorenrolle“ beschrieben werden kann. Im kollaborativen Arbeiten muss zwischen unterschiedlichen Akteuren, Ansichten und Kulturen so vermittelt werden, dass eine gute Zusammenarbeit entstehen kann. Dabei agieren Boundary Spanner als „Filter“ und Übersetzer*in für bestimmte Informationen, als Vertrauensperson in der Kommunikation und als Verhandlungsführer*in bei eventuell auftauchenden Problemen. Empathie, Verhandlungsgeschick und herausragende Fähigkeiten in der Kommunikation und bei Konfliktsituationen zeichnen Boundary Spanner dabei besonders aus. Als Koordinatoren von kollaborativer Arbeit brauchen Boundary Spanner zusätzlich ein tiefes Verständnis und Kenntnis von Praktiken, Abläufen und Zielen in den einzelnen

Akteursgruppen (Williams, 2002; Williams, 2013).

Ähnliche Kompetenzbeschreibungen finden sich auch für die Rolle der Data Stewards, z.B. in Verheul et al. (2019) und Scholtens et al. (2019a) wieder. Data Stewards können dementsprechend als Boundary Spanner angesehen werden, die zwischen den am Datenmanagement beteiligten Akteursgruppen vermitteln.

5.2 Arten von Boundaries

Boundary-Work findet an unterschiedlichen Grenzen statt. Carlile hat drei Arten von Boundaries in Wissenszuständen (*Knowledge Boundaries*) identifiziert, denen jeweils Arten von BOs und Aktivitäten zugeordnet werden, mit denen diese Boundaries überwunden werden können. Diese drei Arten von Boundaries beziehen sich bei Carlile auf Teamarbeit in Unternehmensstrukturen und auf Versuche von *Sensemaking*, also einer Sinnfindung und eine Einordnung von Phänomenen in den jeweiligen individuellen Kontext. BOs können diese Prozesse an den unterschiedlichen Grenzen unterstützen. Es werden folgende Arten von Boundaries beschrieben, die in ihrer Komplexität aufeinander aufbauen (Carlile, 2002; Carlile, 2004, zusammenfassend dargestellt in Tabelle 2):

- **Syntaktische Boundaries** bezeichnen die Grenze, die bei einem *Transfer* von Wissen überschritten werden muss. Hierbei müssen Nachrichten von einem Sender zu einem Empfänger „geschickt“ werden, was eine gemeinsame lexikalische Grundlage für das Verständnis der Nachrichten und des Sendungsprozesses voraussetzt. Diese Grenze wird häufig in der (elektronischen) Verarbeitung von Information angetroffen. Voraussetzung für die Überschreitung der Grenzen ist eine geteilte lexikalische Grundlage und somit eine Form von „*common knowledge*“ (z. B. 1 und 0 bedeuten 1 und 0 in beiden Systemen) als stabile Bedingung für die Informationsübermittlung.
- **Semantische Boundaries** bezeichnen Grenzen, die durch Übersetzungs- und Interpretationsschwierigkeiten zwischen zwei Parteien entstehen. Es liegen unterschiedliche „*worlds of knowledge*“ mit verschiedenen Interpretationen von Begriffen und Konzepten vor, wodurch eine Kommunikation über die Grenze hinweg erschwert wird. Obwohl lexikalische Grundlagen einheitlich können (beide Parteien verwenden das lateinische Alphabet oder sprechen sogar dieselbe Sprache), können Begrifflichkeiten inhaltlich unterschiedlich verstanden werden. An semantischen Boundaries muss *Translation* (Übersetzung) geschehen, damit diese Grenze überwunden werden kann. Es ist wichtig, unterschiedliche Arten der Interpretation zu erkennen und ein gemeinsames Begriffssystem zu entwickeln, gerade wenn mit Objekten, die für beide Parteien eine Neuerung darstellen, gearbeitet werden soll. Hilfreich können hier Standardisierungsprozesse sein.
- **Pragmatische Boundaries** sind Grenzen, die durch unterschiedliche Interessenslagen der Akteure entstehen. Das Transferieren und/oder Übersetzen ihres Wissen über diese Grenze hinweg hat dabei unter Umständen negative Auswirkungen auf die Akteure. Dies liegt daran, dass sich das Wissen auf Probleme innerhalb ihres abgegrenzten

Bereichs bezieht, dass Wissen oft eingebettet ist in konkrete Handlungen und nicht formalisiert ist (implizites statt explizites Wissen) und dass Wissen für die jeweilige Akteursgruppe „von Wert“ ist, der nicht verloren gehen soll. Hierbei kann es sich zum Beispiel um etablierte Praktiken handeln, die für die jeweiligen Akteure „funktionieren“ jedoch bei einer Zusammenarbeit hinderlich wären. Eine *Transformation* des Wissens und die Herausbildung von gemeinsamen Interessen kann dabei helfen, diese Grenzen zu überwinden. Außerdem können Prototypen, wie Zeichnungen oder Diagramme als BOs eingesetzt werden.

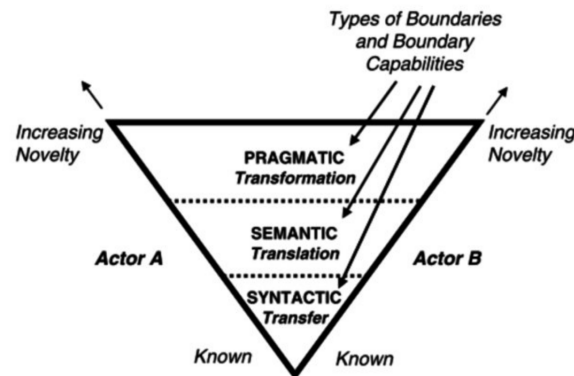


Abbildung 4: Knowledge Boundaries. Aus Carlile, 2004, S. 558

Die von Carlile beschriebenen Arten von Boundaries zeigen sich immer dann, wenn Unterschiede oder Abhängigkeiten zwischen dem jeweiligen Wissensstand der Akteure auffallen. Unterschiede beziehen sich auf die Menge oder die Art des vorhandenen Wissens. Abhängigkeiten von Wissenszuständen bestehen in Verknüpfung mit konkreten Aufgaben, die durch Wissen bewältigt werden können. Einfluss auf die Grenzen hat auch der Neuigkeitswert von Wissen oder Informationen, also wie breit sich neues von bereits vorhandenem Wissen unterscheidet. Je breiter Unterschiede, Abhängigkeiten und Neuigkeitswerte voneinander differieren, desto komplexer werden Wissensgrenzen. Siehe hierzu Abbildung 4. Da die unterschiedlichen Grenzarten aufeinander aufbauen, wird z.B. an pragmatischen Grenzen eine vorherige Überwindung von syntaktischen und semantischen Unterschieden bereits vorausgesetzt. Obwohl Carliles Theorie ein Framework für unterschiedliche Arten von Grenzen in einem recht dichotomen Rahmen vorgibt, also bezogen auf den Unterschied zwischen zwei Akteursgruppen und ohne Betrachtung anderer Stakeholder in den Prozessen, soll im Folgenden die Kategorisierung der einzelnen Grenzarten als Grundlage für die Untersuchung von Grenzen im Umgang mit wissenschaftlichen Daten dienen.

5.3 Verwendung von Boundary Objects

BOs ermöglichen eine Kooperation, der heterogene Ansätze zugrunde liegen, ohne von dieser einen Konsenses zu erfordern. BOs haben im ursprünglichen Konzept nach Star und Griesemer (1989) folgende Eigenschaften (Huvila et al., 2017, S. 2f):

	Syntaktische Boundary	Semantische Boundary	Pragmatische Boundary
Charakter	Voraussetzungen, Einstellungen und Beziehungen zwischen zwei Gruppen sind bekannt. Informationen müssen zwischen den Gruppen übertragen werden.	Es bestehen unterschiedliche Interpretationen und Verständnisse von Konzepten in den Akteurgруппen. Um Informationen übertragen zu können, muss ein gemeinsames Verständnis entwickelt werden.	Es bestehen unterschiedliche Interessen und Motivationen zwischen den Gruppen. Gemeinsame Ziele müssen entwickelt werden.
Aktion	Transferring	Translating	Transforming
Art des BOs	Gemeinsame Repositorien (Speichersysteme)	Standardisierungsverfahren für Methoden und Ergebnisse	Objekte, Modelle, Strategien
Ziel	Schnelle und vollständige Informationsübertragung	Gemeinsames konzeptionelles Verständnis von Informationen, Offenlegung von implizitem Wissen	Veränderung von Interessenslagen und Hemmnissen um Informationen teilbar zu machen.

Tabelle 2: Unterschiedliche Arten von Kognitive Boundaries. Eigene Darstellung, nach Carlile, 2002

- Sie sind (wissenschaftliche) Artefakte, die sowohl in unterschiedlichen Sozialen Welten vorkommen, als auch gleichzeitig das Informationsbedürfnis in jeder Welt befriedigen.
- Sie sind sowohl flexibel genug, um an jeweilige Bedürfnisse und Anforderungen der einzelnen Welten angepasst zu werden, gleichzeitig aber robust genug, um einen übergreifenden Bedeutungskern zu behalten.
- Sie sind in einer gemeinsamen Nutzung schwach strukturiert, werden aber in einer individuellen Nutzung spezifischer strukturiert.
- Sie können abstrakt oder konkret sein.
- Sie können als Übersetzungswerkzeug verwendet werden.

Insbesondere die interpretative Flexibilität mit einem gleichzeitig allgemeinen sinnstiftenden Kern und somit die Möglichkeit, zwischen allgemeinen Bedeutungen und spezifischen Bedeutungen in den jeweiligen Welten wechseln zu können, können als Kerneigenschaften von BOs angesehen werden und machen sie zu Forschungsobjekten in der Wissenschafts- und Technikforschung oder auch den Informationswissenschaften.

Das Konzept der Boundary Objects ist für sich genommen schon recht vage formuliert und kann breit interpretiert werden. Bei einer erneuten Betrachtung des Konzepts und seiner Entwicklung ging Star vor allem auch darauf ein, was kein Boundary Object ist. Der Begriff „boundary“ soll demnach nicht als scharfe „Grenze“ verstanden werden, sondern vielmehr als „*a shared space, where exactly that sense of here and there are confounded.*“ (Star, 2010, S. 602f). Gleichzeitig sind „objects“ nicht gleichzusetzen mit einer unbedingten Materialität. Diese begründet sich erst dadurch, dass „um Objekte herum“ gearbeitet wird und sie somit als gemeinsamer Bezugspunkt verstanden werden.¹⁹ „Scale“ und „Scope“ geben einen Rahmen vor, um BOs zu kategorisieren. *Scale* meint dabei die Reichweite eines Konzepts (oder Objekts), welches als BO definiert wird. Star sieht hierbei das „*organizational level*“ als am besten geeigneter Verwendungsrahmen (ebd., S. 612). *Scope* bezeichnet den Anwendungsrahmen des Konzepts. Hier sind BOs als Forschungsobjekte vor allem dann sinnvoll zu definieren, wenn sie tatsächlich verwendet und mit ihnen in einer interdisziplinären Form „gearbeitet“ wird (ebd., S. 613). BOs „funktionieren“ nicht, wenn die Communities, in denen sie verwendet werden zu groß, zu heterogen oder aber auch zu homogen sind.

In ihrem Artikel charakterisieren Star und Griesemer vier Arten von BOs (Star und Griesemer, 1989):

- **Repositories of things:** geordnete Sammlungen von Objekten.
- **Ideal Types:** idealtypische oder abstrakte Beschreibungen der Realität (z.B. Diagramme).
- **Coincident boundaries:** Objekte, die gleichbleibende Grenzen, aber unterschiedliche Inhalte haben (z.B. der Staat Kalifornien aus territorialer und ideeller Sicht).
- **Standardized Forms:** Formate, die standardisierte Methoden ermöglichen (z.B. Fragebögen).

Diese Arten sind jedoch nicht als abgeschlossene Kategorien anzusehen. Huvila et al. haben

¹⁹Der im deutschen verwendete Begriff „Grenzobjekt“ spiegelt diese nachträgliche Erklärung des Begriffs nicht wider, besser wäre hier die Verwendung des Worts „Schwellenobjekt“ oder sogar „Schwellenkonzept“.

in einer umfassenden Literaturlauswertung zur Verwendung von Boundary Objects in den Informationswissenschaften eine Reihe von weiteren Formen von BOs ausgemacht (Huvila et al., 2017, S. 6):

- **Visionary Objects:** meist abstrakte Begriffe, die gemeinsame Werte und Normen ausdrücken und zu emotionalen Reaktionen führen, z.B. Best Practices oder Codes of Conduct innerhalb eines Unternehmens (nach Briers und Chua, 2001).
- **Primary and secondary BOs:** primary BOs sind Objekte, die im Zentrum aller Aktivitäten stehen (z.B. eine bestimmte Technologie). Secondary BOs werden durch eben diese Aktivitäten geschaffen - sie ermöglichen und sind Ausdruck von kooperativem Handeln, z.B. Dokumentationen zur Verwendung der Technologie (nach Garrety und Badham, 2000).
- **Boundary-objects-in-use/Designated BOs:** Objekte, die bewusst von Akteuren ausgewählt werden, um Grenzen, beispielsweise innerhalb unterschiedlicher Wissenskulturen in einer Organisation zu überwinden, z.B. Strategie-Pläne (nach Levina und Vaast, 2005).

Boundary Objects stellen ein gemeinsames Bedeutungsspektrum oder auch eine Syntax her. Sie fördern die Kommunikation und die Kollaboration unterschiedlicher Akteure oder sozialer Gruppen, indem sie ihnen erlauben, ihr Wissen auf eine Art zu repräsentieren, die sowohl für sie, als auch für andere Akteure Sinn ergibt. Auch können innerhalb des Bedeutungsspektrums von BOs über Grenzen hinweg Zweifel oder Fragen zu einer Idee oder einer Tätigkeit ausgedrückt werden. BOs helfen dabei, unterschiedliche Perspektiven und Wissenskulturen zusammenzubringen und Grenzen dazwischen durch ein gemeinsames Bedeutungsspektrum zu überbrücken, ohne dass die unterschiedlichen Akteure in einen vollständigen Konsens treten müssen (Carlile, 2002).

Grenzobjekte entstehen durch Übersetzungsprozesse auf beiden Seiten der Grenze zwischen Sozialen Welten, die simultan verlaufen und wechselseitig in Beziehung stehen. Durch diese reziproken Beziehungen entsteht an der Grenze ein Raum, in dem unterschiedliche Ansichten und jeweilige Darstellungen der „Wirklichkeit“ aufeinander bezogen werden können, ohne miteinander verschmelzen zu müssen. Die Interpretations- und Übersetzungsprozesse stabilisieren dabei die Beziehungen zwischen den Welten, die jedoch weiterhin voneinander distinkt bleiben (Hörster, Köngeter und Müller, 2013, S. 15). BOs sind an dieser Stelle im Fokus der Übersetzungsprozesse und bilden die Leinwand für die Übersetzungstätigkeit. Gerade bei der noch unsicheren Zusammenarbeit spielen die jeweiligen Akteure, die das BO fortlaufend durch ihre Übersetzungsaktivitäten hervorbringen, eine herausgehobene Rolle. BOs müssen auch immer auf ihren Kontext bezogen werden und befinden sich für eine lange Zeit im Status eines Provisoriums. Um BOs in Kollaborationsprozessen identifizieren zu können, müssen diese daher genauer beobachtet werden, um nachvollziehen zu können, wie sich Problemstrukturen um gemeinsame Objekte gruppieren und wie mit diesen gearbeitet wird (ebd., S.15f). Bezogen auf die von Carlile beschriebenen Grenzarten können BOs an syntaktischen Grenzen durch die Bereitstellung einer gemeinsamen Sprache die Übertragung von Wissen ermöglichen. An semantischen Grenzen können sie als Mittel eingesetzt werden, um Interpretationsmechanismen und Verständigungsprozesse zu spezifizieren. An pragmatischen Grenzen erfüllen BOs die Funktion der Materialisierung von gemeinsamen

Zielen, die durch eine Transformation von bestehenden Wissenszuständen erschaffen werden (Meister, 2011). BOs können an dieser Stelle durch eine Dekontextualisierung des Wissens Differenzen, Perspektiven und Einstellungen in unterschiedlichen Akteursgruppen aufzeigen, ohne dabei persönliche Bezüge herstellen zu müssen (Bechky, 2003).

Das Konzept von BOs hat seit seiner Entwicklung vielfach Anwendung gefunden. In der Forschung zu Design und Innovation wurden unter BOs insbesondere Artefakte als materialisierte, haptische Elemente und Werkzeuge verstanden und in ihrer Integrationsfunktion bei der kooperativen Arbeit beschrieben (z.B. Carlile, 2002). Außerdem wurden BOs bei der Erforschung wissenschaftlicher Zusammenarbeit (Interdisziplinarität) an den Grenzen von unterschiedlichen fachlichen Ansätzen betrachtet (z.B. Fujimura, 1992). Auch in der Kommunikation von Wissenschaft mit Öffentlichkeit (Scientific Communication) können BOs als ein theoretisches Werkzeug dienen (z.B. Polman und Hope, 2014). Nicht zuletzt wird das Konzept in der Forschung zu Organisationstheorien und Wissensmanagement innerhalb von Unternehmen verwendet. Weitere Verwendungsbereiche sind Forschungen zur Lehre und Lernen (z.B. Benn, Edwards und Angus-Leppan, 2013) sowie in den Politikwissenschaften (z.B. Trompette und Vinck, 2009). In den meisten Anwendungsbereichen wird insbesondere die oben erwähnte interpretative Flexibilität in den Vordergrund gestellt. Weniger betrachtet wird ein weiterer Aspekt von BOs, und zwar ihre Relation zu (unsichtbaren) Infrastrukturen, in die sie eingebettet sind und durch die Konventionen, Standards und Normen transportiert werden, sowie ihre Rolle bei der Entstehung von „Wissensartefakten“ wie Klassifikationen, Kategorien, Standards usw. (Huvila et al., 2017 S. 4).

BOs sind keinesfalls als rein statische Objekte anzusehen. Sie interagieren mit sozialen Infrastrukturen, in die sie eingebettet sind, und mit sozialen Identitäten innerhalb der jeweiligen sozialen Gruppen, die sie teilen. Sie fungieren dabei nicht nur als Übersetzungswerkzeug – wie in den ursprünglichen Überlegungen von Star und Griesemer – sondern auch als Vehikel, um soziale Identitäten von Gruppen an ihren Beziehungsgrenzen zu formen. Veränderungen in BOs haben dementsprechend einen Einfluss auf Veränderungen in sozialen Infrastrukturen und Identitäten, die wiederum Veränderungen in angrenzenden Gruppen zur Folge haben können (Gal, Yoo und Boland, 2005).

6 Boundaries im Data Stewardship

6.1 Abgrenzung der betrachteten Akteure

Bisherige Anwendungen von Boundary-Work im Bezug zu Information und Wissen bezogen sich vor allem auf Forschungen in Organisationen oder Teams und ihrer Kollaboration. Dabei sind Grenzen aufgrund von Organisationsstrukturen und -hierarchien leichter festzustellen während im wissenschaftlichen Arbeiten bzw. dem Umgang mit Forschungsdaten Rollenverteilungen häufig nicht konkret geklärt sind. Jedoch besteht auch in wissenschaftlichen Einrichtungen eine organisatorische Trennung zwischen Forschenden, „unterstützenden Tätigen“ (Infrastruktur, Verwaltung) und Entscheidungsträgern, zwischen denen verhaltenskulturelle und strukturelle Differenzen bestehen (Rat für Informationsinfrastrukturen, 2019, S. 4).

Angelehnt an die Darstellung von Stewardship-Rollen in Verheul et al. (2019) und Scholtens et al. (2019b) sowie in Rat für Informationsinfrastrukturen (2019) werden im Folgenden die Akteure aus diesen übergreifenden Kategorien betrachtet: **Forschende, Forschungs- und Informationsinfrastrukturen** (Bibliothek, Rechenzentrum), **Forschungsadministration bzw. Entscheidungsträger** (institutionelle Führungsebene, Verwaltung/Administration). Diese Dreiteilung bietet eine sehr grobe Kategorisierung und kann nicht als abschließende Darstellung der vielen unterschiedlichen Organisationsstrukturen in wissenschaftlichen Einrichtungen angesehen werden.

Unter die soziale Gruppe der Forschenden werden Forschende selbst - jeweils ihren unterschiedlichen Wissenschaftsdisziplinendisziplinen zugeordnet - gefasst. Wissenschaftsdisziplinen entwickeln sich über längere Zeit hinweg. Diese Entwicklungen lassen sich insbesondere seit dem 19. Jahrhundert an institutionalisierten Organisationen (Fachbereiche, Institute), Fachgesellschaften, Ausbildungswegen und zentralen Kommunikationsorganen (z.B. Journals) nachvollziehen. Wissenschaftsdisziplinen bleiben dabei nicht in einer statischen Aufteilung stehen sondern es entwickeln sich im Laufe der Zeit neue Unterdisziplinen, die sich teilweise zu eigenen Disziplinen wandeln (z.B. Bioinformatik), und es entstehen immer wieder interdisziplinäre Forschungsansätze (Stichweh, 1992).

Wissenschaftsdisziplinen können bei der Betrachtung ihrer Aktivitäten und den ihnen zugrundeliegenden Zielen, Normen und Werten auf ihre Forschungskulturen schließen lassen. Im Bezug auf den Umgang mit Forschungsdaten werden bislang in der Regel einzelne Wissenschaftsdisziplinen und die ihnen zugrunde liegenden Forschungskulturen betrachtet, wobei davon ausgegangen wird, dass unterschiedlichen Disziplinen unterschiedliche Forschungskulturen zugrunde liegen und eine Veränderung dieser Kulturen den nachhaltigen und transparenten Umgang mit Forschungsdaten verbessern kann (Borgman, 2012; Borgman, 2015). In der Fachsystematik der DFG werden momentan 213 Fächer gezählt²⁰, die sich den vier großen Wissenschaftsbereichen Geistes- und Sozialwissenschaften, Lebenswissenschaften, Naturwissenschaften und Ingenieurwissenschaften zuordnen lassen. Wenn in dieser Arbeit

²⁰Vgl. https://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtsperiode_2016_2019/fachsystematik_2016-2019_de_grafik.pdf. Abgerufen am 21. Oktober 2019.

von „Forschenden“ als Akteursgruppe mit einer eigenen sozialen Welt die Rede ist, dann sollte dabei immer bedacht werden, dass die Analyse nur auf einer generischen Ebene mit Aktivitäten im Bereich des Datenmanagements erfolgt. Disziplinspezifische Unterschiede in den einzelnen Forschungskulturen sind relevant und verdienen eigenständige Untersuchungen und Forschungsvorhaben, die den Rahmen dieser Arbeit überschreiten würden.

Forschende betreiben sowohl Datenproduktion, als auch Datennutzung und Datenmanagement. Die „Produktion“ von Daten geschieht beispielsweise über Erhebungen, Messungen oder qualitative Methoden, wie die Annotation von Texten oder Bildern. Dabei ist zu beachten, dass die Definition von „Daten“ in den unterschiedlichen Wissenschaftsdisziplinen stark schwanken kann (Borgman, 2012; Borgman, 2015). Forschungsdaten können, je nach Verwendungszweck und Forschungsgegenstand, in unterschiedlichen Formaten und Aggregationszuständen vorliegen. Neben Forschungsergebnissen können auch Code und Software, die die Grundlage der Forschung bilden, als Forschungsdaten angesehen werden. Sowohl quantitative als auch qualitative Daten können Forschungsdaten sein, wobei quantitative Daten häufiger in den Naturwissenschaften und Ingenieurwissenschaften auftreten, während qualitative Daten eher in den Sozial- und Geisteswissenschaften vorkommen.

Datennutzung geschieht während des Forschungsprozesses in Form der Auswertung und der Analyse von Daten. Dabei können Forschende sowohl als Nutzer der eigenen, als auch von bereits erhobenen Daten betrachtet werden, Stichwort Nachnutzung. Das Thema Nachnutzung betrifft eine Vielzahl von Argumenten, die für die Fachdisziplinen und Forschungsformen allerdings auch unterschiedliche Relevanz haben. Die Erhebung und Analyse von Daten ist beispielsweise in Forschungsgebieten, die Großgeräte wie Teleskope benutzen, häufig mit erheblichem zeitlichen und monetären Aufwand verbunden, so dass die Nachnutzung vorhandener Forschungsdaten die Effizienz der Forschung für enorm steigern kann (Piwowar, 2011). Weitere Forschungsfragen können mit Hilfe von bereits vorhandener und nutzbarer Daten beantwortet werden, sofern diese in einem qualitativ hochwertigen und leicht nachzuvollziehenden Format (semantisch und technisch) vorliegen. Neue Forschungsfragen und Perspektiven auf bestehende Annahmen können durch die Nachnutzung und die Kombination von unterschiedlichen Daten auch interdisziplinär entwickelt und beantwortet werden (Rodrigo et al., 2013).

Forschende tragen in der Rolle als Datenmanager insbesondere Verantwortung für die Aufbereitung der eigenen Daten angefangen mit „Rohdaten“ bis hin zu kontextualisierten Forschungsdaten. Dabei stehen Datenbereinigung, Datenbeschreibung und andere Formen der Dokumentation im Vordergrund. Forderungen nach Open Science und FAIR-konformen Daten erfordern hierbei einen erhöhten Aufwand, besonders bei der Beschreibung der erhobenen Daten mit Metadaten und bei der Dokumentationsarbeit.

Unter Forschungs- und Informationsinfrastrukturen fallen in etwa jene Bereiche, die die Europäische Kommission als „Data Services“ beschreibt (Hodson et al., 2018). Informationsinfrastrukturen können als ein Netzwerk von grundlegenden Ressourcen, die das wissenschaftliche Arbeiten ermöglichen definiert werden (Bowker et al., 2009, S. 98). Sie existieren häufig im Hintergrund der Forschung und werden nur dann sichtbar, wenn ihre Funktion gestört ist oder ihre Verstetigung, z.B. nach einer Projektphase, fehlt. Im Sinne der Wissenschafts- und

Technikforschung sind Informationsinfrastrukturen Netzwerke aus Maschinen und Menschen mit technologischen, politischen, organisatorischen Komponenten. Infrastrukturen entstehen aus menschlichem Handeln und beeinflussen dieses wiederum. Sie sind eingebettet in größere (soziale) Strukturen von Einrichtungen und Communities. Dieser Sicht auf Infrastrukturen als Ökosysteme liegt folgendes Definitionsschema zugrunde (Star und Ruhleder, 1996, S. 113):

- **Embeddedness:** Infrastruktur ist in andere Strukturen, sowohl soziale als auch technologische, eingebettet.
- **Transparency:** Infrastruktur ist zwar nicht immer sichtbar, muss jedoch nicht für jede Aufgabe neu erschaffen werden.
- **Reach:** Infrastruktur ist temporär oder räumlich nicht einmalig und wird nicht nur für ein bestimmtes Ereignis verwendet.
- **Learned as part of membership:** Die Bedeutung von Infrastruktur ist innerhalb einer sozialen Gruppe definiert. Andere Gruppen nehmen Infrastrukturen unter Umständen anders wahr.
- **Links with conventions of practice:** Infrastruktur beeinflusst und ist beeinflusst durch die Art der Tätigkeiten in einer Gruppe.
- **Embodiment of standards:** Infrastrukturen können mit anderen Infrastrukturen oder Werkzeugen in einem standardisierten Format verbunden werden.
- **Built on installed base:** Infrastruktur baut auf vorhandenen Basisstrukturen auf und „überbaut“ sie in ihrer Weiterentwicklung.
- **Becomes visible upon breakdown:** Infrastruktur wird vor allem erst dann sichtbar, wenn ihre Funktion gestört ist.

Entsprechend werden Informationsinfrastrukturen in dieser Arbeit als die technologischen Produkte einerseits (Repositorien, Datenbanken, Tools für die Datenverarbeitung...) und die Tätigkeiten der Menschen in den Infrastruktureinrichtungen (Bibliothekar*innen, Informatiker*innen, Administrator*innen...) andererseits angesehen.

Infrastruktureinrichtungen werden bestimmte Aufgaben des Datenmanagements zuteil. Sie übernehmen die Daten von den Forschenden und stellen die Kuratierung, Veröffentlichung und Langzeitverfügbarkeit sicher. Außerdem stellen sie für Forschende Services und Software zur Datenerhebung und -auswertung zur Verfügung. Dafür müssen Geschäfts- und Betriebskostenmodelle entwickelt und angewendet werden. Eine dritte Funktion der Infrastruktureinrichtungen liegt im Bereich der Schulung und Beratung von Akteuren zum Datenmanagement. Schulungsangebote zum Thema Datenmanagement sind häufig an Bibliotheken (im Sinne der Informationskompetenzvermittlung) angesiedelt. Infrastruktureinrichtungen können durch kompetente Strategieentwicklungen nachhaltig zur Qualität des Datenmanagements auf Seiten der Forschenden beitragen. Zudem müssen sie, als technologischer Grundbaustein für das Datenmanagement, Anforderungen an Kosten und personelle Ressourcen verhandeln (Hartmann, Jacob und Weiß, 2019).

Als Entscheidungsträger werden Akteure aus der Forschungsadministration und der Leitungsebene definiert. Entscheidungsträger sind dafür verantwortlich, Leitlinien (Policies) für das Datenmanagement zu entwickeln und Anreize zu verstärken, zum Beispiel durch Förderent-

scheidungen im Sinne von Open Science oder neuen Reputationsmechanismen. Policies sollten dabei, wenn möglich, institutionsübergreifend aufeinander abgestimmt sein und im engen Kontakt mit allen anderen Akteuren entwickelt werden (Neylon, 2017). Zudem fallen Entscheidungsträgern Aufgaben im Bereich der Mittelvergabe, der Haushaltsplanung sowie der (wissenschaftspolitischen) Steuerung des Wissenschaftssystems zu. Auf institutioneller Ebene müssen Entscheidungsträger hierbei ihre eigenen Interessen (z.B. nach Wirtschaftlichkeit) bewahren, sollten jedoch in enger Abstimmung mit den anderen Akteursgruppen stehen.

Hochschulgovernance im Sinne des „*new public management*“ (NPM) stellt durch inter- und intra-organisatorischen Steuerungsstrukturen den Rahmen für die Governance der Forschung selber her. Allgemein bestehen aber folgende Regulationskategorien:

- Staatliche Regulierung der Hochschulen vor allem in Bezug auf Personal- und Ressourcenensatz sowie zu Rahmen- und Prüfungsordnungen. Hiermit sind vor allem die Bildungsministerien der einzelnen Bundesländer gemeint, aber auch das Bundesministerium für Bildung und Forschung (BMBF).
- Außenregulierung durch den Staat oder andere Akteure, wie die Hochschulrektorenkonferenz (HRK) oder den Wissenschaftsrat (WR) in Form von Zweck- oder Zielvereinbarungen.
- Akademische Selbstorganisation in Gremien wie Fachbereichsräte und Senate.
- Selbststeuerung der Hochschulen durch hierarchisch organisierte Leitungsorgane wie Präsidien und Dekanate.

In Deutschland hat die „Freiheit der Wissenschaft“ durch Autonomiebestreben auf Ebene der akademischen Selbstorganisation auf der einen Seite und die Außenregulierung durch wissenschaftspolitische Akteure auf der anderen Seite eine lange Tradition (Schimank, 2010; Schimank, 2015). Betrachtet werden in der folgenden Analyse Entscheidungsträger auf institutioneller Ebene in der akademischen Selbstorganisation und der Leitungsorgane von Hochschulen. Diese handeln jedoch auch im Kontext staatlicher Vorgaben sowie (disziplinärer) Positionen von Fachgesellschaften und anderer, z.B. rechtlicher Rahmenbedingungen.

Entscheidungsträger verhalten sich auch im Sinne des automatisierten Forschungsmonitorings (z.B. über Forschungsinformationssysteme) als Konsumenten von Daten. Digitales, datenbasiertes Berichtswesen ist dabei im NPM eine Möglichkeit, um die Effizienz von Wissenschaftsarbeit auf quantitativer Ebene zu messen. Außerdem geben sie Rahmenbedingungen und Anforderungen für das Datenmanagement vor.

6.2 Data Stewardship Tätigkeiten an Boundaries

Anhand der Ergebnisse des LRCDM-Reports (Verheul et al., 2019) lässt sich ablesen, dass für ein erfolgreiches Data Stewardship im konzeptionellen Sinne alle oben beschriebenen Akteure beteiligt sind. Die Aufgaben und Tätigkeiten, die in diesem Report beschrieben sind, unterscheiden zwischen drei personalisierten Rollen, den „embedded Data Stewards“, „generic

and advisory Data Stewards” und „policy, strategy and advisory Data Stewards”. Während diese Unterteilung im Hinblick auf Aufgabenverteilung und Verantwortungsübernahme durch konkrete Personen in wissenschaftlichen Einrichtungen durchaus sinnvoll ist, sollen in dieser Arbeit jedoch zunächst alle Tätigkeiten im Bezug auf das umfassende Konzept Data Stewardship betrachtet werden. Hierfür eignen sich die Ergebnisse des Projektes EOSCpilot zu Kompetenzen und Aufgaben im Data Stewardship (Whyte et al., 2018) eher, da Tätigkeitsbereiche²¹ anhand des Data Lifecycles und darüber hinaus beschrieben werden und so eine umfassendere Sicht auf das Konzept Data Stewardship erlauben. In der folgenden Analyse wird davon ausgegangen, dass das Konzept „Data Stewardship” von einer oder mehreren Personen in der Rolle des Data Stewards vollzogen wird. Anhand von konzeptionellen Aufgaben und den an ihnen beteiligten Akteuren können so spezifische Tätigkeiten an der Überschreitung von (kognitiven) Boundaries in der Kollaboration zwischen den unterschiedlichen Akteuren anhand der Kategorisierung von Carlile (2004) abgeleitet werden.

Einschränkend muss erwähnt werden, dass die aufgeführten Tätigkeiten einem theoretischen und konzeptionellen Rahmen entstammen, während der LRCDM-Report sich auf Ergebnisse aus Auswertungen von Stellenausschreibungen, Umfragen und Interviews bezieht und somit einen (kleinen) Ausschnitt aus der konkreten Arbeitsrealität an niederländischen Forschungseinrichtungen zeigt. Zudem beziehen sich die von Carlile entwickelten Kategorien speziell auf Knowledge Boundaries. Das in dieser Arbeit entwickelte theoretische Konzept geht jedoch davon aus, dass eben diese in der Kooperation unterschiedlicher Akteure im Data Stewardship auftauchen und dass Data Stewards hier den größten Teil ihrer Arbeit verrichten.

Im EOSCpilot werden 59 Tätigkeiten als „*relevant to professional groups involved in open science data stewardship*” (Whyte et al., 2018, S. 37) beschrieben. Im Projekt wurden die Akteure Forschende, Data Scientist, Data Service Engineer und Data Manager/Curator jeweils in ihren Rollen als Nutzende oder Bereitstellende von Services betrachtet. Wie bereits in Kapitel „Abgrenzung der betrachteten Akteure” herausgearbeitet spielen jedoch neben Infrastruktur (im EOSCpilot in etwa Data Service Engineer und Data Manager) und Forschenden (im EOSCpilot Researcher und Data Scientist/-analyst) auch Akteure aus der Forschungsadministration, die sogenannten „Policymaker” oder Entscheidungsträger eine Rolle im Data Stewardship. Diese Akteure werden im EOSCpilot nicht dezidiert betrachtet. Sie werden in dieser Arbeit berücksichtigt, da sich neuste Verwendungen des Begriffs Data Stewardship (bspw. in Verheul et al., 2019) auch auf die Planungsebene von Seiten der Entscheidungsträger beziehen. Die Tätigkeiten untergliedern sich in folgende Abschnitte:

- **Plan and design:** Planung des Forschungsvorhabens und des zugehörigen Forschungsdatenmanagements, beispielsweise im Bezug auf verwendete Metadatenschemata und Datenmodelle sowie Softwarelösungen.
- **Capture and process:** Erfassung der Daten während der Forschungsaktivität und ihre Weiterverarbeitung, mit dem Fokus auf Datenorganisation, Workflowmanagement und Softwareverwendung.
- **Integrate and analyse:** Datenintegration und -analyse, einschließlich mathemati-

²¹Im Bericht „*competences*” genannt, allerdings werden eher Tätigkeiten als Kompetenzen beschrieben.

schen, kreativen und theoretischen Analyseverfahren, Datenabfragefunktionalitäten und integrative Datenmodellierung.

- **Appraise and preserve:** Datenbewertung und -übernahme für die Langzeitarchivierung, beispielsweise durch Qualitätsprüfung, Datentransfer, (Format-)Migration und rechtlicher Prüfung.
- **Publish and release:** Veröffentlichung von Datensätzen inklusive z.B. Dokumentation, Zugriffskontrollen und Lizenzierungsmethoden.
- **Expose and discover:** Optimierung der veröffentlichten Datensätze für die Auffindbarkeit und Nachnutzbarkeit, beispielsweise durch die Verwendung von kontrollierten Vokabularen, Sicherstellung der Zitierbarkeit oder Visualisierung.
- **Govern and assess:** Steuerung und Bewertung von Forschungsaktivitäten im Bezug auf Strategieentwicklung, FAIR-konforme Daten, Sicherheitsmanagement und Data Governance.
- **Scope and resource:** Reichweite und Ressourcenmanagement für das Datenmanagement mit Hinblick z.B. auf Services, Change Management oder anfallende Kosten.
- **Advise and enable:** Beratung und Befähigung zum Datenmanagement unter anderem durch Trainings, Ausbildung oder Kollaboration.

Nicht alle Akteure sind gleich stark in die unterschiedlichen Aktivitäten eingebunden. Die folgende Auswertung zeigt in vereinfachter Form, wo welche Akteure beteiligt sind und welche Art von Grenzarbeit in Frage kommt, beruhend auf der Zuordnung der Tätigkeit zu den Grenzkategorien nach Carlile. Die beteiligten Akteure sind farblich in Infrastruktureinrichtungen (**rot**), Entscheidungsträger (**grün**) und Forschende (**blau**) unterschieden. Bei einer Beteiligung an der jeweiligen Tätigkeit sind die entsprechenden Felder farblich markiert. Anhang A zeigt die Zuordnung der Tätigkeiten (dargestellt als gelber Kreis) in einer schematischen Darstellung an den Schnittstellen zwischen den unterschiedlichen „sozialen Welten“ der Akteure. Hieraus soll ersichtlich werden, an welcher Position innerhalb der einzelnen Sphären die jeweiligen Tätigkeiten angesiedelt sind.

Plan and design

Die **Planung des Datenmanagements** erfolgt meistens mit Hilfe eines Datenmanagementplans und involviert alle drei beteiligten Parteien. Dabei stehen insbesondere die Infrastruktureinrichtungen und die Forschenden direkt in Kontakt, da diese häufig bei der Erstellung von DMPs eng zusammen arbeiten. Die Entscheidungsträger sind jedoch genauso stark, wenn auch eher indirekt, involviert, da Anforderungen (Policies) von dieser Seite direkten Einfluss auf die Planung des Datenmanagements haben können. Grenzarbeit muss hier insbesondere in der Übersetzung dieser Anforderungen sowohl für Forschende als auch für Infrastrukturanbieter, die ihre Services entsprechend ausrichten sollten, geleistet werden. Es handelt sich dementsprechend um eine semantisch gelagerte Boundary, an der alle drei Akteure beteiligt sind und die *Übersetzungstätigkeit* erfordert.

Die **Spezifikationen von Datenmodellen** und das **Design von Datenbanken** für die Speicherung von (Forschungs-)Daten beteiligen insbesondere die Infrastruktureinrichtungen,

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**

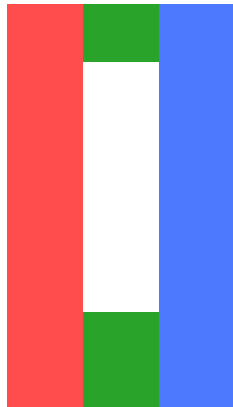
Tätigkeit	Art der Grenzarbeit	Beteiligte
Planung des Datenmanagements (DMP)	Übersetzungstätigkeit	
Open Data Modelle und Datenbankdesign	Übertragungstätigkeit / Übersetzungstätigkeit	
Metadaten und PID Spezifizierung	Übersetzungstätigkeit	
Anforderungen an Open Source Software und Services	Übertragungstätigkeit / Übersetzungstätigkeit	
Bewertung von Repositorien und Datenmanagement-Plattformen	Übertragungstätigkeit	

Tabelle 3: Tätigkeiten „Plan and design“, siehe Anhang A.1

jedoch in sehr enger Zusammenarbeit mit den Forschenden. Datenmodelle sollten so gestaltet sein, dass sie die anfallenden Daten verständlich abbilden können. Hierfür müssen Daten in Datenbanken so vorliegen, dass auf sie von unterschiedlichen Systemen aus zugegriffen werden kann. Gleichzeitig müssen die Modelle entsprechende disziplinspezifische Anforderungen widerspiegeln. Datenmodelle müssen demnach auf einer semantischen Ebene anfallende Forschungsdaten verständlich abbilden. Auch bei dieser Art von Grenzarbeit wird sowohl an syntaktischen als auch an semantischen Boundaries gewirkt. Es handelt sich sowohl um eine *Übertragungstätigkeit* als auch um eine *Übersetzungstätigkeit* zwischen Forschenden und Infrastruktureinrichtungen.

Die **Spezifizierung von Metadatenmodellen und Persistent Identifiern** findet vor allem auf Ebene der Metadatenschemata in enger Zusammenarbeit zwischen Forschenden und Infrastruktureinrichtungen statt. Metadatenschemata werden traditionell von Akteuren der Infrastruktur im Austausch mit Expert*innen aus den jeweiligen Fachdisziplinen erarbeitet. Dabei sollten sie, ähnlich wie Datenmodelle, die Struktur der in den jeweiligen Disziplinen anfallenden Daten abdecken können. Im Sinne von FAIR-konformen Forschungsdaten sollten Metadatenschemata zusätzlich so gestaltet sein, dass eine optimale Auffindbarkeit und Interoperabilität gewährleistet werden kann. Hierbei hilft die Verwendung von PID-Systemen, wobei geklärt werden muss, in welcher Granularität Daten oder Datensätze mit PIDs beschrieben werden sollten. Ein gemeinsames Verständnis von Datenstrukturen wird hierbei für beide Akteursgruppen vorausgesetzt. Es lässt sich eine semantische Boundary erkennen, die eine *Übersetzungstätigkeit* nach sich zieht.

Im Sinne einer offenen Wissenschaft sollten möglichst nicht-proprietäre Softwarelösungen (Open Source Software) und Services Verwendung finden. Die **Verhandlung der Anforderungen** an diese Systeme geschieht zwischen Forschenden und Infrastruktureinrichtungen. Dabei müssen spezifischen Anforderungen entsprechend an die Infrastruktureinrichtung kommuniziert werden. Außerdem müssen auch rückwirkend eventuell schon vorhandene

Angebote publik gemacht oder unter Umständen angepasst werden. Dieser Prozess deutet auf eine semantische Boundary hin, die *Übersetzungstätigkeit* erfordert. Gleichzeitig ist die Verwendung von Open Source Software in manchen Forschungsprozessen noch nicht etabliert. Hier müsste ein Veränderungsprozess angestrebt werden, der sowohl die Infrastruktureinrichtungen (Anbieterseite) als auch die Forschenden (Nutzende) betrifft. An dieser Stelle zeigt sich demnach zusätzlich ein Beispiel für eine pragmatische Boundary, die die Notwendigkeit von *Transformation* mit sich bringt.

An der **Bewertung von Repositorien und Datenmanagement-Plattformen** sind vor allem Forschende und Infrastruktureinrichtungen, in geringerem Maße auch Entscheidungsträger beteiligt. Repositorien und Plattformen werden von Seiten der Infrastruktureinrichtungen zur Verfügung gestellt und/oder betreut und müssen den Anforderungen der Forschenden entsprechen und gleichzeitig innerhalb von rechtlichen und datenschutztechnischen Rahmenbedingungen liegen. Diese Anforderungen müssen für eine Bewertung entsprechend übersetzt werden, um auftauchende semantische Boundaries überwinden zu können. An dieser Stelle kann dann von *Übersetzungstätigkeit* gesprochen werden.

Capture and process

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**





Tätigkeit	Art der Grenzarbeit	Beteiligte	
Bereitstellung des Workflows und Dokumentation	Übersetzungstätigkeit		
Datenbankmanagement	X		
Prototyping von Software	X		
Datenerhebung und Nachnutzung von Open Data	Übersetzungstätigkeit / Transformation		
Dateibenennungen und -organisation	Übertragungstätigkeit		
Datenprovenienz und Software Versionierung	Übertragungstätigkeit		

Tabelle 4: Tätigkeiten „Capture and process“, siehe Anhang A.2

Workflows und Dokumentation zu Datenerhebungen werden vor allen von Seiten der Forschenden konzipiert, jedoch in Tools und Services von Seiten der Infrastruktureinrichtungen abgebildet. Die Bereitstellung dieser Werkzeuge setzt die Beteiligung der Infrastruktureinrichtungen voraus. Dies zieht gleichzeitig eine Abstimmung über Prozesse und Art der Datenerhebung nach sich. Bei dieser Abstimmung müssen die Prozesse zwischen Forschenden und Infrastruktureinrichtungen übersetzt werden - es handelt sich um eine semantische Boundary an der *Übersetzungstätigkeit* geschehen muss.

Das **Management von Datenbanken**, die aufgrund solcher Prozesse entstanden sind, geschieht jedoch dann ausschließlich auf der Seite der Infrastruktureinrichtungen. Während Konzeption und Modellierung von Datenbanken und Datenmodellen eine strategische Arbeit bedeutet, ist das Datenbankmanagement eine eher operativ geartete Tätigkeit, die wenig Kollaboration erfordert. Aus diesem Grund bilden die Infrastruktureinrichtungen an dieser Stelle die einzige beteiligte Akteursgruppe und die Tätigkeit wird an keiner Boundary verortet. Ähnlich verhält es sich mit dem **Prototyping von Software**, was als überwiegend infrastrukturelle Aufgabe angesehen wird. Auch hier wird keine Art von Grenzarbeit vermutet.

An der **Datenerhebung** und der eventuellen **Nachnutzung von offenen Daten** sind insbesondere die Forschenden beteiligt, die die Datenerhebung durchführen sowie die Infrastruktureinrichtungen, die (teilweise) die Werkzeuge zur Verfügung stellen. Besonders aber wenn es um die Nachnutzung bereits vorhandener Daten geht, müssen Infrastruktureinrichtungen und Forschende kollaborieren. Infrastruktureinrichtungen können dabei Daten auffindbar und nachnutzbar machen. Auch Entscheidungsträger sind, wenn auch nicht sehr stark, in den Prozess von Datenerhebung involviert, dafür aber umso mehr in die Nachnutzung. Entsprechende Policies für ein korrektes Vorgehen und zur Transparenz von Datenerhebungen sowie Anerkennungsverfahren und Anreize zur Nachnutzung können hierbei Einfluss auf das Forschungsvorhaben nehmen. Es besteht dabei eine semantische Boundary in Hinblick auf Anforderungen und Tätigkeiten bei der Erhebung und Nachnutzung. Gleichzeitig zeigt sich auch eine pragmatische Boundary da die Nachnutzung von bereits erhobenen Daten bisher in vielen Forschungsdiziplinen sowohl von Seiten der Entscheidungsträger als auch der Forschenden noch nicht in bestehende Forschungskulturen übergegangen ist. Die Art der Grenzarbeit, die an dieser Stelle geleistet werden muss ist demnach sowohl eine *Übersetzungsarbeit* als auch eine *Transformation*.

Dateibenennungen und die **Organisation von Dateien** sind vor allem Aufgaben der Datenerhebenden, also der Forschenden. Jedoch sind auch im Hinblick auf die spätere Übernahme in ein zentrales Datenmanagementsystem die Akteure von Seiten der Infrastruktureinrichtungen beteiligt. Daten sollten daher so strukturiert sein, dass sie einfach in ein neues System übertragen werden können. Hierfür muss die Organisation auf beiden Seiten verständlich sein. Normalerweise sollten dabei keine größeren Übersetzungstätigkeiten anfallen, daher kann diese Art der syntaktischen Boundary mit Hilfe von *Übertragungstätigkeit* überwunden werden. Etwa gleich geartet sind auch Tätigkeiten bei der **Beschreibung der Provenienz von Daten** und **Software Versionierung**. An dieser Stelle müssen ebenfalls Informationen in einer leicht zu transferierenden Art vorliegen. An dieser syntaktischen Boundary wird dies durch *Übertragungstätigkeiten* möglich.

Integrate and analyse

Die **Anwendung mathematischer und statistischer Verfahren** auf Daten ist eine Tätigkeit, die eher der Gruppe von Forschenden zuzuschreiben ist, da diese die Datenauswertung vornehmen. Es wird daher keine Art der Grenzarbeit identifiziert. Jedoch liegt die Tätigkeit

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**

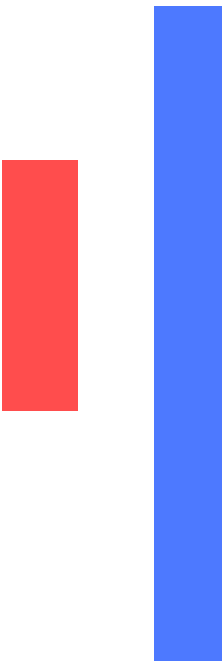
Tätigkeit	Art der Grenzarbeit	Beteiligte
Anwendung mathematischer und statistischer Verfahren	X	
Kritisches Denken und Theorie-Bildung	X	
Kreative Problemlösung, Flexibilität	Übertragungstätigkeit / Übersetzungstätigkeit	
Entwicklung von Open Source Software und Services	Übersetzungstätigkeit	
Datentransformation und -integration	Übertragungstätigkeit	
Datamining, Datenanfrage und -interpretation	X	
Vorrausschauende Modellierung und Analyse	X	
Methoden im Maschinellen Lernen	X	

Tabelle 5: Tätigkeiten „Integrate and analyse“, siehe Anhang A.3

der Infrastruktureinrichtungen recht nah, da verwendete Methoden auch in der späteren Datenübernahme beschrieben werden bzw. nachträglich erneut durchführbar sein sollten. Dies würde für eine syntaktische Boundary und Übertragungsarbeit sprechen, wird jedoch hier aufgrund der bislang noch im Wandel befindlichen Tätigkeitsmerkmale in der Infrastruktur (noch) nicht so klassifiziert. Ebenfalls sind **kritisches Denken und Theorie-Bildung** auf Grundlage von Datenanalysen Forschenden zuzuschreiben. Auch an dieser Stelle wird ebenfalls eher keine Grenzarbeit zwischen den hier beschriebenen Akteuren (allerdings natürlich zwischen Akteuren innerhalb der Gruppe der Forschenden) vermutet.

Kreative Problemlösung und Flexibilität ist, genau wie kritisches Denken, eher eine Kompetenz als eine Tätigkeit. Sie soll vor allem Forschenden zugeschrieben werden, betrifft jedoch auch Akteure der Infrastruktureinrichtungen, die entweder an Problemlösungsprozessen beteiligt sind oder diese später in Datenmodellen abbilden sollten. Im ersten Fall müssen Lösungsansätze und unterschiedliche Problemsichten über eine semantische Boundary hinweg übersetzt und verstanden werden, wobei *Übersetzungstätigkeit* helfen kann. Im zweiten Fall müssen Lösungswege so übertragen werden, dass sie entsprechend abgebildet werden können. *Übertragungstätigkeiten* werden an dieser eher syntaktischen Art von Boundary verlangt.

Die **Entwicklung von Open Source Software und Services** findet vor allem auf Seite der Infrastruktureinrichtungen statt, die jedoch in einem Austausch mit den Forschenden stehen sollte um deren Anforderungen umsetzen zu können. Hierbei müssen erneut Anforderungen entsprechend über eine semantische Boundary transportiert und bei der Entwicklung bedacht

werden. *Übersetzungstätigkeit* sollte an dieser Stelle die Grenzarbeit dominieren.

Tätigkeiten im Bereich der **Datentransformation und -integration** betreffen sowohl Akteure aus dem Bereich der Infrastruktureinrichtungen als auch der Forschenden. Die Tätigkeiten können sowohl innerhalb der einzelnen Akteursgruppe als auch in Kollaboration geschehen. Bei der hier betrachteten Kollaboration müssen Daten so vorliegen, dass sie über eine syntaktische Grenze hinweg verständlich und verwendbar sind. Entscheidungen zur Transformation und Integration können darauf basierend getroffen werden, wenn beispielsweise auf beiden Seiten dasselbe Metadatenschema verwendet wird. Dies zeigt ein Beispiel für eine *Übertragungstätigkeit* in der Grenzarbeit.

Datamining, Datenabfrage und -interpretation, Datenmodellierung und -analyse sowie **Methoden des Maschinenlernens** werden letztendlich nur den Forschenden zugeordnet, wobei sich sicherlich bei allen Tätigkeiten diskutieren ließe, ob und inwieweit die Infrastruktureinrichtungen beteiligt ist. Dies betrifft vor allem Technologien im Machine Learning oder Data Mining. In diesem Fall wird allerdings keine Art von Grenzarbeit definiert.

Appraise and preserve

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**

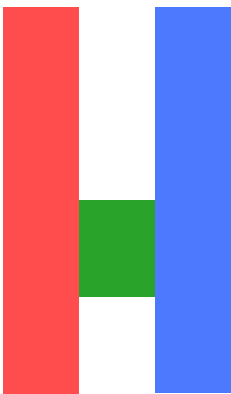
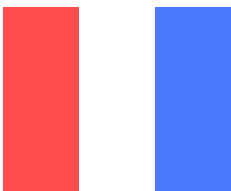




Tätigkeit	Art der Grenzarbeit	Beteiligte
Datenaufbereitung und -dokumentation für die Nachnutzung	Übersetzungstätigkeit	
Qualitätsmanagement von Daten und Sicherstellung offener Standards	Übersetzungstätigkeit / Transformation	
Einhaltung ethischer und rechtlicher Vorgaben	Übersetzungstätigkeit / Transformation	
Datenübertragung und Langzeitvorhaltung	Übertragungstätigkeit	
Migration von Formaten und Datenträgern	X	
Evaluation von Software und Softwarearchivierung	X	

Tabelle 6: Tätigkeiten „Appraise and preserve“, siehe Anhang A.4

Die **Datenaufbereitung und -dokumentation für die Nachnutzung** wird in Kollaboration zwischen Forschenden und Infrastruktureinrichtungen durchgeführt. Dabei sollten Forschende die entstandenen Daten bereits von Beginn des Forschungsvorhabens an so

aufbereiten und dokumentieren, dass sie ohne zusätzlichen Aufwand veröffentlicht, verstanden und nachgenutzt werden können. Die Akteure der Infrastruktureinrichtungen wirken in dieser Tätigkeit unterstützend. Anforderungen an Datendokumentationen, interoperable Standards usw. müssen zwischen Infrastruktureinrichtungen und Forschenden über eine semantische Boundary hinweg transferiert werden. Grenzarbeit wird hier in Form von *Übersetzungstätigkeit* geleistet.

Für das **Qualitätsmanagement** von Daten besonders im Bezug auf die **Verwendung von offenen Standards** müssen die Akteure der Infrastruktureinrichtungen ebenfalls wieder mit Akteuren von Seiten der Forschenden zusammenarbeiten. Die Hauptarbeit liegt hier bei den Infrastruktureinrichtungen, jedoch müssten Standards und Datenqualität bereits bei der Erhebung der Daten berücksichtigt werden. Qualitätsmerkmale und zu verwendende Standards müssen einerseits an die Forschenden über eine semantische Boundary hinweg kommuniziert werden, andererseits kann es sein, dass in manchen Forschungscommunities offene Standards bislang nicht verwendet wurden. Ist dies der Fall, muss zusätzlich eine pragmatische Boundary überwunden und Überzeugungsarbeit geleistet werden. Entsprechend kann an dieser Stelle sowohl *Übersetzungsarbeit* als auch *Transformation* als Form der Grenzarbeit hilfreich sein.

Entscheidungsträger verabschieden Leitlinien für den korrekten Umgang mit Daten. Die **Einhaltung der ethischen und rechtlichen Vorgaben** kann nur dann in Kollaboration mit den Forschenden und den Infrastruktureinrichtungen geschehen, wenn diese Vorgaben sowohl der Forschungsrealität entsprechen als auch von allen Akteuren verstanden werden. Es handelt sich hier also um eine semantische Boundary in Form von einem gemeinsamen Verständnis einerseits und andererseits auch zu einem Teil um eine pragmatische Boundary, da entsprechende Vorgaben konkrete Forschungsaktivitäten nachhaltig beeinflussen können und diese wiederum Einfluss auf die Vorgaben haben können. Grenzarbeit an dieser Stelle besteht also aus einem Großteil von *Übersetzungstätigkeit* aber auch einem Anteil *Transformation*.

Die **Datenübertragung und Langzeitvorhaltung** findet insbesondere auf Seite der Infrastruktureinrichtungen statt, jedoch muss zumindest bei der Übertragung der Daten mit Forschenden kollaboriert werden. Dabei muss sichergestellt werden, dass die Daten entsprechend aufbereitet und im richtigen Format für eine Übertragung vorliegen und in das zentrale Speichersystem (z.B. ein Repositorium) übernommen werden können. An dieser syntaktischen Boundary muss eine *Übertragungstätigkeit* vollzogen werden.

Für die Langzeitarchivierung muss unter Umständen eine **Migration von Daten und Datenträgern** stattfinden. Auch für die verwendete Software sollte eine **Softwarearchivierung** durchgeführt werden. Diese Tätigkeiten finden so gut wie ausschließlich auf Seite der Infrastruktureinrichtungen statt. Im Bezug auf Softwarebewertung und -archivierung muss dabei eventuell mit den Forschenden enger zusammengearbeitet werden, als bei der Datenmigration, da Software aus eigens für das Forschungsvorhaben erstellten Bausteinen bestehen kann. Jedoch wird zunächst für keine der Tätigkeiten eine Form von Grenzarbeit identifiziert.

Publish and release

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**

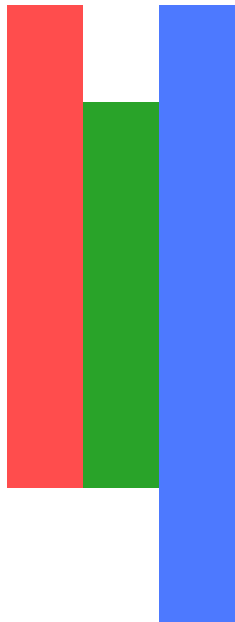
Tätigkeit	Art der Grenzarbeit	Beteiligte
Dokumentation für öffentliche Nutzung und Verständnis	Übersetzungstätigkeit / Transformation	
Management von Zugangskontrolle	Übertragungstätigkeit / Übersetzungstätigkeit	
Ethische Anwendung von Patenten und Lizenzen	Übersetzungstätigkeit / Transformation	
Open Access Publikation und Selbstarchivierung	Übersetzungstätigkeit / Transformation	
Teilen über offene Repositorien oder Plattformen	Transformation	
Teilnahme an der Entwicklung offener Lösungen außerhalb des wissenschaftlichen Kontextes	X	

Tabelle 7: Tätigkeiten „Publish and release“, siehe Anhang A.5

Die **Dokumentation für die öffentliche Nutzung und Verständnis** bezieht sich auf eine Form von Dokumentation der Daten, die auch von einem nicht-wissenschaftlichen Publikum oder von Forschenden aus anderen Disziplinen verstanden werden kann. Hierbei sollten Forschende und Infrastruktureinrichtungen eng zusammenarbeiten. Anforderungen an diese Art der Dokumentation müssen einerseits über eine semantische Boundary hinweg von beiden Seiten verstanden werden. Andererseits kann diese Tätigkeit einen zusätzlichen Aufwand für beide Akteure bedeuten, was eine Veränderung von etablierten Aktivitäten auf beiden Seiten nach sich ziehen kann und eine pragmatische Boundary signalisiert. Die Grenzarbeit an dieser Stelle ist also sowohl geprägt von *Übersetzungstätigkeit* als auch von *Transformation*.

Für das **Management von Zugangskontrollen** müssen Entscheidungsträger, Infrastruktureinrichtungen und Forschende zusammenarbeiten. Entscheidungsträger legen dabei die rechtlichen und ethischen Rahmenbedingungen fest, aufgrund derer die Services zur Zugangskontrolle von den Infrastruktureinrichtungen angepasst werden müssen. Forschende müssen bei der Aufbereitung und Übergabe der Daten diese Rahmenbedingungen beachten und sicherstellen, dass auch nur die Daten veröffentlicht werden, die keine sensiblen Inhalte anzeigen oder die entsprechend anonymisiert wurden. Die Anforderungen an die Zugangskontrolle müssen einerseits allen Akteuren bekannt sein (syntaktische Boundary) und andererseits auch verstanden werden (semantische Boundary). Hierfür muss sowohl eine *Übertragungstätigkeit* als auch eine *Übersetzungstätigkeit* angestrebt werden.

Die **ethische Anwendung von Lizenzen und Patenten** betrifft, ähnlich wie die Zugangskontrolle, alle drei Akteursgruppen. Auch hier müssen Rahmenbedingungen definiert, verstanden und auf entsprechende Services und Tätigkeiten angewendet werden. Insbesondere die Lizenzierung von Daten für die Veröffentlichung muss im Hinblick auf eine angestrebte Nachnutzbarkeit gut bedacht werden. Entscheidungsträger müssen hierbei die Anforderungen klar kommunizieren, Akteure der Infrastruktureinrichtungen sollten dafür Sorge tragen, dass technische Systeme diese Anforderungen entsprechend abbilden können und Forschende sollten dazu angehalten werden, ihre Daten so zu lizenzieren, dass eine Nachnutzbarkeit ohne größere Hürden möglich ist. Entsprechend können sowohl semantische als auch pragmatische Boundaries identifiziert werden. *Übersetzungstätigkeit* und *Transformation* werden bei der Grenzarbeit vorausgesetzt.

Open Access Publikation und Selbstarchivierung erfordern ein hohes Maß an Zusammenarbeit zwischen Infrastruktureinrichtungen, Entscheidungsträgern und Forschenden. Hierbei geht es vor allem um eine gesteigerte Akzeptanz und Sensibilisierung für OA-Publikationen auf Seiten der Forschenden, um die Entwicklung eines Anreizsystems durch die Entscheidungsträger (z.B. konkrete Forderungen, Forschungsergebnisse in OA zu veröffentlichen oder Vorteilen wie der Vereinfachung des Reportings) und die Bereitstellung von Beratungsangeboten und Services auf Seiten der Infrastruktureinrichtungen. Gleichzeitig sollten Forschende auf ihr Recht zur Zweitveröffentlichung und auf die unterschiedlichen OA-Modelle aufmerksam gemacht werden. Es handelt sich hierbei um eine recht komplexe Art von Boundary, die sowohl aus semantischen als auch aus pragmatischen Anteilen besteht, da sowohl ein gemeinsames Verständnis als auch eine Veränderung im Publikationsverhalten hergestellt werden sollte. *Übersetzungstätigkeiten* und *Transformation* sind dabei unabdingbar.

Noch komplexer gestalten sich Tätigkeiten zum **Teilen von Daten auf offenen Repositorien oder Plattformen**. Teilweise besteht vor der Hürde der Veröffentlichung der Forschungsergebnisse in OA bereits eine ablehnende Haltung gegenüber dem Teilen von Forschungsdaten auf Seiten der Forschenden. Auch hier sind Entscheidungsträger und Infrastruktureinrichtungen durch das Schaffen von Anreizen, Rahmenbedingungen, Strukturen und Services eingebunden. Obwohl natürlich auch eine Notwendigkeit zur Übersetzung von Konzepten und Rahmenbedingungen für das Teilen von Daten besteht, wird diese Tätigkeit an einer überwiegend pragmatische Boundary verortet, an der in erster Linie *Transformation* stattfinden muss. Wenn Forschende außerhalb des wissenschaftlichen Kontextes an der **Entwicklung von offenen Lösungen** teilnehmen, bezieht diese Tätigkeit zwar andere Akteure aus der Gruppe der Forschenden ein, fokussiert sich jedoch eher auf diese Gruppe. Daher kann bei dieser Tätigkeit keine Grenzarbeit identifiziert werden.

Expose and discover

Die **Anwendung kontrollierter Vokabulare oder Ontologien** betrifft sowohl die Infrastruktureinrichtungen als auch die Forschenden. Innerhalb der Infrastruktureinrichtungen werden üblicherweise kontrollierte Vokabulare entwickelt und für die Anwendung bereitgestellt.

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**











Tätigkeit	Art der Grenzarbeit	Beteiligte
Anwendung kontrollierter Vokabulare/Ontologien	Übersetzungstätigkeit	 
Bereitstellung von Metadaten und PIDs	X	
Visualisierung und Präsentation von Forschungsergebnissen	Übersetzungstätigkeit	 
Evaluation von Repositorien und Publikationsplattformen	Übersetzungstätigkeit	 
Suchen in Repositorien und Datenbanken	Übersetzungstätigkeit	 
Zitieren von Forschungsergebnissen	Übertragungstätigkeit	 

Tabelle 8: Tätigkeiten „Expose and discover“, siehe Anhang A.6

Diese Vokabulare müssen jedoch wiederum auf die Art der Forschungsergebnisse anwendbar sein, bzw. direkt verwendet werden. Aus diesem Grund muss zwischen den Akteursgruppen ein gemeinsames Verständnis über die Art der Vokabulare, ihre Einsatzmöglichkeiten und ihren Nutzen hergestellt werden. Es handelt sich um eine semantische Boundary, an der *Übersetzungstätigkeiten* geleistet werden.

Die tatsächliche **Bereitstellung von Metadaten und Persistent Identifiern** ist Aufgabe der Infrastruktureinrichtungen und auch nur dort verortet. Eine Grenzarbeit kann im Bezug auf diese konkrete Tätigkeit nicht identifiziert werden.

Bei der **Visualisierung und Präsentation von Forschungsergebnissen** müssen Infrastruktureinrichtungen und Forschende auf eine Art und Weise kollaborieren, die eine integrale Repräsentation der Ergebnisse ermöglicht. Wenn Forschende ihre Ergebnisse nicht selber visualisieren, sollten sie ihre Ergebnisse so beschreiben, dass die Infrastruktureinrichtungen als Anbieter für Werkzeuge zur Visualisierung oder anderer Arten von Repräsentation die in den Daten enthaltenen Ergebnisse verständlich und sachgerecht wiedergeben kann. An dieser semantischen Grenze besteht also die Notwendigkeit für eine *Übersetzungstätigkeit*.

Um eine möglichst gute Auffindbarkeit und Nutzbarkeit der Forschungsergebnisse zu gewährleisten sollte vor der Veröffentlichung eine **Evaluation von Repositorien und Publikationsplattformen** stattfinden. Beteiligt hieran sind die Infrastruktureinrichtungen, wobei jedoch auf disziplinübliche Anforderungen und Konventionen der Forschenden Rücksicht genommen werden muss und gleichzeitig gewonnene Ergebnisse auf verständliche Art und Weise an die Forschenden übergeben werden sollten. Um dies zu erreichen kann es zu semantischen Boundaries kommen, die mit Hilfe von *Übersetzungstätigkeiten* überwunden werden können.

Die **Suche in Repositorien und Datenbanken** findet in Kollaboration zwischen der Infrastruktureinrichtungen als Anbieter der Services und Forschenden als Nutzende der Services statt. Dabei sollten die Infrastruktureinrichtungen Anforderungen und Suchverhalten von Forschenden kennen und bei der Entwicklung von Services berücksichtigen. Diese Anforderungen müssen an die Infrastruktureinrichtungen in verständlicher Form übertragen werden. Außerdem sollten Akteure der Infrastruktureinrichtungen Suchfunktionalitäten so präsentieren, dass sie für Forschende verständlich und verwendbar sind. *Übersetzungstätigkeiten* charakterisieren diese semantische Boundary.

Das **Zitieren von Forschungsausgaben** ist vor allem Aufgabe der Forschenden. Jedoch ist auch an einem Teil auch die Infrastruktureinrichtungen beteiligt, da dort Services für ein einfaches Zitieren (z.B. über PIDs) oder das Nachvollziehen von Zitationen zur Verfügung gestellt werden. Die Tätigkeit des Zitierens sollten entsprechend über eine syntaktische Boundary hinweg übertragbar sein. *Übertragungstätigkeiten* sind hier als Grenzarbeit angemessen.

Govern and assess

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**




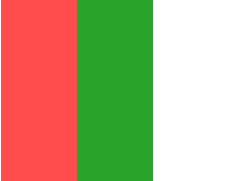



Tätigkeit	Art der Grenzarbeit	Beteiligte
Offene Forschungsstrategie / -vision	Übersetzungstätigkeit / Transformation	
Verbreitung der FAIR Prinzipien und Open Access Leitlinie	Übersetzungstätigkeit / Transformation	
Forschungsintegrität, Sensibilisierung für Anerkennung und Impact	Übersetzungstätigkeit / Transformation	
Informationssicherheit und Risikomanagement	Übertragungstätigkeit / Übersetzungstätigkeit	
Data Governance und Verwendung von Daten von Drittanbietern	Übertragungstätigkeit / Übersetzungstätigkeit	
Sicherheitsmanagement für Speichertechnologien	X	
Beteiligung an Qualitätssicherung und Peer-Review	X	

Tabelle 9: Tätigkeiten „Govern and assess“, siehe Anhang A.7

Entscheidungsträger und Forschende sind maßgeblich beteiligt an der **Entwicklung einer offenen Forschungsstrategie/-vision**. Hierbei müssen beide Akteursgruppen Strategien entwickeln, die sich aufeinander abstimmen lassen. Entscheidungsträger können durch

Leitlinien z.B. zum Umgang mit Forschungsdaten und zu OA-Publikationen die Forschungskulturen „top-down“ beeinflussen, während Entwicklungen innerhalb der Forschungskulturen zu neuen Ausrichtungen innerhalb der Strategieentwicklungen von Entscheidungsträgern führen können. Diese Abstimmungsverfahren sind geprägt durch die Entwicklung von einem gemeinsamen Verständnis von offener Wissenschaft aber auch durch gemeinsame Zielsetzungen und angestrebte Ergebnisse. Es handelt sich hierbei sowohl um eine semantische als auch, und sogar überwiegend, um eine pragmatische Boundary, an der *Übersetzungstätigkeiten* und *Transformation* als Grenzarbeit sehr wichtig sind.

Im gleiche Zuge sind die **Verbreitung der FAIR Prinzipien und OA-Leitlinien** geprägt von einer Zusammenarbeit zwischen Entscheidungsträgern und Forschenden aber auch von der Infrastruktureinrichtungen, die besonders im Hinblick auf die FAIRifizierung des Forschungsprozesses und seinen Ergebnissen eine tragende Rolle spielen. Die Umsetzung der FAIR Prinzipien und von OA-Leitlinien setzt ein gemeinsames Verständnis von Anforderungen, die sich aus ihnen ergeben, voraus. Zudem müssen, gerade im Hinblick auf FAIR-konforme Daten, Infrastruktureinrichtungen und Forschende, aber auch Entscheidungsträger die Prozesse, Rahmenbedingungen und Services überarbeiten und teilweise umstellen. Auch hier bestehen sowohl semantische als auch pragmatische Boundaries, mit der Folge von benötigter *Übersetzungsarbeit* sowie *Transformation*.

Forschungsintegrität und Sensibilisierung für Reputationsverfahren und Impact betreffen Forschende und Entscheidungsträger. Auf beiden Seiten muss zusammengearbeitet werden, damit die Integrität und Bewertung von Forschungsleistungen einer offenen Wissenschaft entsprechen kann. Dabei müssen sowohl Anforderungen von beiden Seiten übersetzt als auch neue Ziele entwickelt werden. Semantische und pragmatische Boundaries können auftreten und mit *Übersetzungstätigkeiten* und *Transformation* überwunden werden.

Akteure aus den Infrastruktureinrichtungen kollaborieren bei Tätigkeiten im Bereich **Informationssicherheit und Risikomanagement** miteinander. Dabei müssen Anforderungen von Entscheidungsträgern auf entsprechende Sicherheitssysteme abgebildet werden, die durch die Infrastruktureinrichtungen entwickelt werden. Eine Übertragung von Anforderungen über eine syntaktische Grenze kann bei einfacheren Entscheidungen zu Zugriff und Sicherheit geschehen, während kompliziertere Verfahren und Prozesse im Risikomanagement eher semantische Boundaries überwinden müssen. Grenzarbeit an dieser Stelle kann also sowohl als *Übertragungstätigkeit* als auch *Übersetzungstätigkeit* nötig sein.

Die **Data Governance und die Verwendung von Daten von Drittanbietern** betreffen, ähnlich wie die Informationssicherheit, die Infrastruktureinrichtungen und die Entscheidungsträger. Entscheidungsträger geben dabei Leitlinien und Regeln vor, die durch die Infrastruktureinrichtungen umgesetzt werden sollen. Auch hier können Regeln leicht verständlich sein oder eine Übersetzung benötigen. Wieder finden sich syntaktische und semantische Boundaries und Grenzarbeit in Form von *Übertragungstätigkeit* und *Übersetzungstätigkeit*.

Keine Arten von Grenzarbeiten können für die Tätigkeiten **Sicherheitsmanagement von Speichertechnologien** und **Beteiligung an Qualitätsmanagement (Peer-Review)** iden-

tifiziert werden. Speichertechnologien werden ausschließlich von Seite der Infrastruktureinrichtungen betreut. Ebenso sind Forschende (bisher) als einzige der hier betrachteten Akteure an Peer-Review-Verfahren in der Wissenschaft beteiligt.

Scope and ressource

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**









Tätigkeit	Art der Grenzarbeit	Beteiligte
Finanzierung von Open Science	Transformation	
Ausdauernde Tätigkeit in vielfältiger offener Wissenschaft	X	
Service Level Management	Übersetzungstätigkeit	
Change Management	Transformation	
Bereitstellung von Workflows und Management von Provenienz	X	
Management von Cloud- und Speichertechnologien	X	
Authentifizierungs- und Authorisierungsmanagement	X	
Kosten von Datenmanagement und Datenarchivierung	Übersetzungstätigkeit / Transformation	

Tabelle 10: Tätigkeiten „Scope and ressource“, siehe Anhang A.8

Bei der **Finanzierung von Open Science** mit Fördergeldern müssen Entscheidungsträger und Forschende aufeinander abgestimmt zusammenarbeiten um eine offene Wissenschaft voranzutreiben. Entscheidungsträger sollten dabei Finanzierungsentscheidungen von Kriterien der Open Science abhängig machen während Forschende ihre Vorhaben entsprechend vorschlagen sollten. Jedoch müssen hierfür zunächst gemeinsame Ziele erarbeitet werden um eine solche Abstimmung zu ermöglichen. Es handelt sich um eine pragmatische Boundary, bei deren Auftreten *Transformation* als Grenzarbeit geleistet werden muss.

Eine **ausdauernde Tätigkeit in der offenen Wissenschaft** liegt bei den Forschenden selbst. Zwar können Anreize von Seiten der Entscheidungsträger und verlässliche Infrastruktureinrichtungen diese Tätigkeit beeinflussen, jedoch wird keine explizite Grenzarbeit identifiziert.

Beim **Service Level Management** geht es auf der Seite der Akteure der Infrastruktureinrichtungen darum, die angebotenen Services aktuell gemäß den Anforderungen und

Rahmenbedingungen von Seiten der Forschenden und der Entscheidungsträger zu halten und ggf. anzupassen. Hierbei müssen die Anforderungen entsprechend auf Services übertragen werden was eine semantische Boundary andeutet. *Übersetzungstätigkeit* kann an dieser Stelle dabei helfen, diese Herausforderungen zu meistern.

Change Management liegt vor allem in der Verantwortung der Entscheidungsträger und betrifft die Infrastruktureinrichtungen und die Forschenden. Für ein kompetentes Change Management sollten alle Akteure kollaborieren und Prozesse und Ziele miteinander abgeglichen werden. Veränderungen erfordern an einer pragmatischen Grenze hierbei Grenzarbeit in Form von *Transformation*.

Workflows und Provenienzmanagement, Management von Cloud- und Speichertechnologien sowie von **Authentifizierungs- und Authorisierungsverfahren** liegen in der Verantwortung der Infrastruktureinrichtungen. Natürlich sind diese Tätigkeiten auch eingebettet in die vorgegebenen Rahmenbedingungen und reagieren in Teilen auf Anforderungen von Seiten der Forschenden, jedoch können keine tatsächlichen Formen der Grenzarbeit identifiziert werden.

An der Berechnung der **Kosten des Datenmanagements und der -archivierung** sind alle drei Akteure beteiligt. Kosten entstehen sowohl auf Seite der Forschenden, die Daten erheben, als auch insbesondere auf Seite der Infrastruktureinrichtungen, die das spätere Datenmanagement und die -archivierung betreiben. Anforderungen von beiden Seiten müssen von den Entscheidungsträgern verstanden und bewilligt werden. *Übersetzungstätigkeiten* als Form der Grenzarbeit stehen an dieser semantischen Boundary im Vordergrund, obwohl es auch dazu kommen kann, dass auf einer pragmatischen Ebene Boundaries überwunden werden müssen, was eine *Transformation* voraussetzt.

Advise and enable

Der **Aufbau offener interdisziplinärer Kollaborationen** betrifft zum größten Teil Akteure von Seiten der Forschenden. Auf einer kleinteiligeren Ebene besteht hier sicherlich Bedarf nach Grenzarbeit zwischen unterschiedlichen Forschungskulturen, aber für diese Analyse kann keine Art der Grenzarbeit identifiziert werden.

Alle hier genannten Akteure sollten sich im Zuge einer offenen Wissenschaft an der **Zusammenarbeit mit zukünftigen Nutzern der Forschung** beteiligen. Diese Tätigkeit betrifft die Entwicklung von zugänglichen und offenen Infrastrukturen, das Schaffen von administrativen Rahmenbedingungen, die eine offene Wissenschaft ermöglichen und eine transparente und nachvollziehbare Forschungsarbeit. Dabei müssen Anforderungen von außen (den „Stakeholdern“) aber auch aus den einzelnen Akteursgruppen übersetzt werden um eine Kollaboration zu ermöglichen. An dieser semantischen Grenze kann durch *Übersetzungstätigkeit* gewirkt werden.

Die **Entwicklung eines offenen Forschungsprofils** ist vor allem geprägt durch die Zu-

Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**

Tätigkeit	Art der Grenzarbeit	Beteiligte		
Aufbau offener interdisziplinärer Kollaborationen	X			
Zusammenarbeit mit Verwendern der Forschung und anderen Stakeholdern	Übersetzungstätigkeit			
Entwicklung eines offenen Forschungsprofils	Transformation			
Training für offene Forschungsmethoden und Services	Übersetzungstätigkeit			
Beitrag zur Ausbildung und Kompetenzentwicklung	Transformation			
Beitrag zu offenen Research and Education Spaces (RES), Netzwerken und Standards	X			
Supervision und Mentoring	X			

Tabelle 11: Tätigkeiten „Advise and enable“, siehe Anhang A.9

sammenarbeit von Entscheidungsträgern und Forschenden. Auch bei dieser Tätigkeit müssen Rahmenbedingungen und Forschungsaktivitäten aufeinander abgestimmt werden, so dass ein Profil entwickelt werden kann. Dabei reicht simple Übersetzung häufig nicht aus, weswegen an dieser Stelle eine pragmatische Boundary identifiziert werden kann, die *Transformation* voraussetzt.

In Kollaboration zwischen Infrastruktureinrichtungen und Forschenden kann ein **Training für offene Forschungsmethoden und Services** durchgeführt werden. Hierbei müssen Desiderate der Forschenden den Infrastruktureinrichtungen bewusst werden. Außerdem müssen Trainings entwickelt werden, die verständlich sind und auch eventuelle nicht artikulierten Wissenslücken auf Seiten der Forschenden füllen können. *Übersetzungstätigkeiten* und Vermittlung zwischen den Akteuren sind an dieser semantischen Boundary besonders gefordert.

Über Trainings hinaus kann auch von Seiten der Entscheidungsträger ein **Beitrag zur Ausbildung und zur Kompetenzentwicklung** geleistet werden, zum Beispiel indem Gelder und weitere Angebote für Forschende bereitgestellt werden. Auch hierbei sollte auf Desiderate reagiert werden, jedoch müssen auch Ziele und Vorhaben aufeinander abgestimmt werden. Daher wird an dieser Stelle eher eine pragmatische Boundary gesehen, an der Stelle Grenzarbeitende in Form von *Transformation* wirken können.

An der **Entwicklung von neuen Standards und Netzwerken** beteiligen sich eher Akteure aus Infrastruktureinrichtungen während **Supervision und Mentoringtätigkeit** relativ singulär innerhalb der Akteursgruppe der Forschenden durchgeführt wird. In dieser Analyse konnten an beiden Stellen keine Formen der Grenzarbeit identifiziert werden, da diese wohl eher auf einer granulareren Ebene (innerhalb derselben Akteursgruppe) betrachtet werden müsste.

7 Data Stewardship als Boundary Work

7.1 Akteure und Boundaries

Akteursgruppen im Data Stewardship

Die Analyse zeigt, dass Tätigkeiten im Data Stewardship sehr häufig mindestens zwei der drei hier betrachteten Akteure involvieren. Aktivitäten, die nur eine der drei Akteursgruppen betreffen finden sich auf Seite der Infrastruktureinrichtungen und der Forschenden, jedoch nie auf Seiten der Entscheidungsträger.

Die Infrastruktureinrichtungen sind in der Analyse an 44 von 59 Tätigkeiten involviert. Bei den Schritten „Plan and design“, „Capture and process“ und „Expose and discover“ sind die Infrastruktureinrichtungen an allen Aktivitäten beteiligt, was auf die infrastrukturellen Kerntätigkeiten im Data Stewardship schließen lässt. Dabei arbeiten die Infrastruktureinrichtungen besonders häufig mit den Forschenden zusammen. Es konnten nur zwei Tätigkeiten erkannt werden, an denen Infrastruktureinrichtungen und Entscheidungsträger ohne Beteiligung der Forschenden kollaborieren. Diese sind *Informationssicherheit und Risikomanagement* sowie *Data Governance und die Verwendung von Daten von Drittanbietern*. Hierbei wird davon ausgegangen, dass Entscheidungen auf administrative Ebene Einfluss auf die Entwicklung von entsprechenden Infrastrukturen haben und Forschende an den Entscheidungen kaum oder gar nicht beteiligt sind.

Akteur	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Anzahl	44	21	47

Tabelle 12: Akteure und Anzahl der Tätigkeiten mit Beteiligung

Die Entscheidungsträger sind in der Analyse an 21 von 59 Aktivitäten beteiligt. Auffällig ist, dass für die Entscheidungsträger keine Tätigkeit identifiziert werden konnte, die nur diese Gruppe betrifft. Entscheidungsträger wurden in der Analyse stets in Kollaboration mit Forschenden, Infrastruktureinrichtungen oder beiden Gruppen den Tätigkeiten zugeordnet. Dies kann durchaus daran liegen, dass „Data Stewardship“ im EOSCpilot anhand von sehr praktischen Tätigkeiten, die operativ gesehen „nah an den Daten liegen“ beschrieben ist. Dennoch zeigt die Analyse, dass es durchaus Tätigkeiten gibt, die enge Kollaboration mit den Entscheidungsträgern voraussetzen. Diese finden insbesondere im Bereich *Govern and assess* und *Publish and release* statt. Vor allem Data Governance, welches laut EOSCpilot dem Data Stewardship zugeordnet werden kann, betrifft die Entscheidungsträger und die Policy-Ebene direkt. Policies haben aber auch Einfluss auf die Veröffentlichung von Daten und anderen Forschungsergebnissen, die im Sinne von Open Science am besten Open Access publiziert werden sollten. Hierbei befinden sich die Akteure der Entscheidungsträger in einer „Top-Down“-Position und können Anreize schaffen und Hemmnisse beseitigen.

Forschende sind an 47 von 59 Aktivitäten beteiligt und sind somit die Akteursgruppe mit den

meisten Beteiligungen. Als Datenproduzierende und -nutzende Akteure kann den Forschenden im Zusammenarbeit mit Infrastruktureinrichtungen die aktivste Rolle im Data Stewardship zugeschrieben werden. Dies liegt auch daran, dass Data Stewardship sich über den gesamten Forschungsprozess hinweg zieht und in vielen Fällen die Infrastruktureinrichtungen an primär auswertenden und analysierenden Prozessen nicht direkt eingebunden werden. Ein Beispiel hierfür ist der Bereich *Integrate and analyse*, der deutlich den Forschenden zugeordnet werden kann. Hierbei sind besonders die Anwendung mathematischer und statistischer Verfahren, kritisches Denken und Theorie-Bildung und Analysetätigkeiten in der Domäne der Forschenden zu verankern. Sobald der Umgang mit Daten jedoch den rein wissenschaftlichen Kontext verlässt, ist eine Kollaboration zwischen Forschenden und Infrastruktureinrichtungen die tragende Ebene des Data Stewardships.

Boundaries im Data Stewardship

Boundaries und zugehörige Arten von Grenzarbeit konnten in der Analyse immer dann identifiziert werden, wenn mindestens zwei Akteure an einer Tätigkeit beteiligt waren. Insgesamt konnten 5 rein syntaktische Boundaries, 5 syntaktische Boundaries mit semantischen Anteilen, 14 rein semantische Boundaries, 10 semantische Boundaries mit pragmatischen Anteilen und 5 rein pragmatische Boundaries definiert werden (siehe Tabelle 13). Die Auswertung zeigt, dass der Fokus von Data Stewardship auf Übersetzungstätigkeiten liegt, Übertragungstätigkeiten auch unterstützt werden müssen und pragmatische Boundaries nur an wenigen Stellen zu überwinden sind. Bei der Auswertung ist jedoch zusätzlich zu beachten, dass die Arten der Boundaries nicht distinkt voneinander sind, sondern aufeinander aufbauen und es sich um Abstufungen im Bezug auf in den Akteursgruppen jeweils divergierende Wissens- und Informationsverwendung handelt. Für diese Analyse wurde davon ausgegangen, dass sich Wissen in Praktiken widerspiegelt und Tätigkeiten daher auf Knowledge Boundaries hindeuten bzw. sich diese in Tätigkeiten zeigen können.

Syntaktische Boundaries finden sich vor allem an Stellen, an denen konkret mit den Daten gearbeitet wird und zugehörige Praktiken beiden Akteursgruppen weitestgehend bekannt sein sollten. Dies betrifft zum Beispiel die Dateibenennung, bei der als gemeinsame lexikalische Grundlage, das Benennungsschema, von beiden Akteursgruppen gleichermaßen verwendet werden sollte oder die Überführung von Daten in ein anderes System, wobei beiden Akteursgruppen die Systemstruktur bekannt sein sollte. Versionierungen von Software und Daten sowie Datenprovenienz und das Zitieren von Daten können ebenfalls als einfache Übertragungstätigkeit angesehen werden. Die Gruppe der Entscheidungsträger ist an keiner reinen Übertragungstätigkeit beteiligt. Dies deutet darauf hin, dass syntaktische und „einfach“ zu überwindende Boundaries zwar zwischen Infrastruktureinrichtungen und Forschenden vorkommen, Interaktionen mit Entscheidungsträgern aber immer mindestens einen Anteil von Übersetzungstätigkeiten beinhalten müssen. Beispiele hierfür sind das Management von Zugangskontrollen oder weitere Rechte- und Sicherheitsmanagementaspekte. An diesen Stellen können Policies und administrative Anforderungen nicht einfach übertragen werden, sondern sollten zusätzlich so übersetzt werden, dass alle Akteure dasselbe Verständnis dieser Praktiken haben.

Kategorie	Anzahl
syntaktisch	5
syntaktisch/semantisch	5
semantisch	14
semantisch/pragmatisch	10
pragmatisch	5

Tabelle 13: Kategorien von Boundaries und identifizierte Anzahlen

Dieser Umstand deutet schon an, dass semantische Boundaries im Data Stewardship klar überwiegen. Der Großteil der Tätigkeiten im Data Stewardship verlangt die Übersetzung von Begriffen, Konzepten und Ansichten zwischen den Akteursgruppen um eine Kollaboration herstellen zu können. Dies wird besonders deutlich, wenn Anforderungen (z.B. an Services oder an Software/Tools) formuliert werden müssen oder wenn Evaluationen von Qualitätsmerkmalen oder Services durchgeführt werden. Semantische Boundaries zeigen sich auch immer dann, wenn zwei Akteure (meist die Infrastruktureinrichtungen und die Forschenden) direkt mit den Daten arbeiten z.B. in der Visualisierung, Datenaufbereitung, durch die Verwendung von Standards und/oder kontrollierten Vokabularen sowie in der Dokumentation. Dies kann unter anderem daran liegen, dass „Daten“ und zugehörige Begriffe und Konzepte in den jeweiligen Welten der Akteursgruppen unterschiedliche Bedeutungen und Stellenwerte haben können. Übersetzungstätigkeiten sind Grundvoraussetzung um diese Boundaries zu überwinden.

Gleichzeitig zeigt sich an vielen semantischen Grenzen auch eine Tendenz zu einer pragmatischen Grenze, bei der reine Übersetzungstätigkeiten nicht mehr ausreichen. Stattdessen müssen Einstellungen und Ziele der Akteursgruppen geändert werden um eine Kollaboration zu erreichen. Beispiele hierfür sind Tätigkeiten, die sich auf die wissenschaftliche Arbeit selbst auswirken können, wie Inhalte des Open Science Movements (inklusive Open Access Transformation und Open/FAIR Data), Veränderungen in der (freien) Lizenzierung oder Qualitätsstandards. An reinen pragmatischen Grenzen sind immer Entscheidungsträger beteiligt, was ihre Rolle als „Top-Down“-Akteure in Transformationsprozessen nochmals bestätigt. Die Tätigkeiten *Offenes Teilen von Daten*, *Finanzierung von Open Science*, *Change Management*, *Entwicklung offener Forschungsprofile* und *Beiträge zu Kompetenzentwicklungen* wurden reinen pragmatischen Boundaries zugeordnet. Diese Aktivitäten richten sich in die Zukunft und adressieren die Nachhaltigkeit von Data Stewardship Prozessen. An pragmatischen Grenzen findet Grenzarbeit demnach mit einem Fokus auf nachhaltige Strategien und Verstetigungen statt.

Zusammenfassend führt die Auswertung der Analyse zu folgender These:

These 1. Data Stewardship findet an den Grenzen zwischen unterschiedlichen sozialen Welten statt und erfordert dabei die Zusammenarbeit unterschiedlicher Akteure, sowohl in der Übertragung von Informationen als auch insbesondere für die Herstellung von gemeinsamen Verständnisgrundlagen und einer (Neu-)Ausrichtung von Werten und Zielen für ein nachhaltiges Datenmanagement.

Hiermit sind zuerst das *Konzept* Data Stewardship sowie praktische Tätigkeiten im Data Stewardship gemeint. Die Auswertung bezieht sich auf eine theoretische Auflistung von Tätigkeiten und wurde anhand einer abstrakten Betrachtung durchgeführt. Jedoch können verschiedene Akteursgruppen aus unterschiedlichen sozialen Welten für einen Großteil der Tätigkeiten identifiziert werden, die miteinander kollaborieren müssen. Hier zeigt die Auswertung, dass Übersetzungstätigkeiten für die meisten gelisteten Aktivitäten Voraussetzungen sind und dass Nachhaltigkeitsaspekte eine Transformation innerhalb der jeweiligen sozialen Welt mit sich bringen.

7.2 Data Stewards als Boundary Spanner

Boundary-Work findet an den Schwellen zwischen unterschiedlichen sozialen Welten und ihren Akteuren statt. Für das Data Stewardship als Sammlung von Praktiken wurden diese Akteure und die Arten der Grenzen, an denen sie aufeinandertreffen, in Kapitel 6 beschrieben. Kollaboration zwischen den Akteuren und Welten als eine Voraussetzung für ein erfolgreiches Data Stewardship ist eine Grundannahme dieser Arbeit. Diese Kollaboration kann durch die Arbeit von designierten Data Stewards als Personen, die an den Grenzen als Kommunikator*innen/Interpret*innen und Koordinator*innen (nach Williams, 2013) eingesetzt werden, gefördert werden. Data Stewards erfüllen dann die Rolle von Boundary Spannern und vermitteln zwischen den unterschiedlichen Akteursgruppen. Besonders aktiv geschieht dies an Stellen, an denen mehrere Akteursgruppen beteiligt sind.

Die Analyse zeigt eine Reihe von Aktivitäten, die alle drei betrachteten Akteursgruppen involvieren. Es handelt sich hierbei um:

- Die Planung des Datenmanagements und die Erstellung eines Datenmanagementplans
- Die Bewertung von Repositorien und Datenmanagement-Plattformen
- Die Datenerhebung und die Nachnutzung von Open Data
- Das Management von Zugangskontrollen auf Daten
- Die ethisch korrekte Anwendung von Patenten und Lizenzen
- Open Access Publikationen und die Selbstarchivierung von Forschungsergebnissen
- Das Teilen von Forschungsdaten über offene Repositorien/Plattformen
- Die Verbreitung der FAIR Data Prinzipien und einer Open Access Policy
- Service Level Managemenet
- Change Management
- Die Berechnung und Bewertung von Kosten für das Datenmanagement und die Langzeitarchivierung

- Die Zusammenarbeit mit (zukünftigen) Nutzer*innen der Forschungsausgaben sowie anderen Stakeholdern

Diese Tätigkeiten lassen sich grob in zwei Sparten unterteilen, die auch Rückschlüsse auf die Tätigkeiten der Data Stewards zulassen. Zum einen handelt es sich um Planungs- bzw. Managementaufgaben (DMPs, Evaluation von Repositorien, Zugangskontrollen, Service Level Management, Kostenberechnung) und zum anderen um strategische Aufgaben, bei denen unterschiedliche Praktiken zusammengeführt werden müssen und die auf eine Veränderung der „Wissenschaftskultur“ in Richtung Open Science hinsteuern (Nachnutzung (offener) Daten, Lizenzierungsfragen, Open Access und das Teilen von Forschungsdaten, FAIR Data, Change Management, Nutzbarkeit).

Im Hinblick auf Planungs- oder Managementaufgaben können Data Stewards als Koordinatoren die Aktivitäten zwischen den Akteuren abstimmen und entsprechend orchestrieren. Beispielsweise können Anforderungen übersetzt und zusammengeführt werden. Sie können alle Seiten dabei unterstützen, entsprechende Aufgabenbereiche und Verantwortlichkeiten zuzuordnen und operative Aspekte des Datenmanagements entsprechend kollaborativ zu planen. Bei diesen Aufgaben handelt es sich meist um Übertragungs- oder Übersetzungstätigkeiten, bei denen Data Stewards als vermittelnde Bindeglieder und Boundary Spanner wirken können. Aus semantischer Sicht ist hierbei das Schaffen eines gemeinsamen Vokabulars und ein übergreifendes Verständnis von Anforderungen die Voraussetzung für eine gelungene Kollaboration. Data Stewards als Boundary Spanner müssen dafür intensiv mit Praktiken, Sichtweisen und Anforderungen der einzelnen Akteursgruppen vertraut sein und diese entsprechend übermitteln können.

Bei Tätigkeiten, die eine Veränderung (Transformation) in der wissenschaftlichen Arbeit bedeuten können, wirken Data Stewards in erster Linie als Kommunikatoren und Mediatoren zwischen den einzelnen Akteursgruppen. Hierbei handelt es sich häufig um Situationen, in denen unterschiedliche Zielvorstellungen und abweichende, in den Akteursgruppen aber manifestierte Arbeitsweisen und Formen des impliziten Wissens zusammengebracht werden müssen. Beispiele hierfür sind u.a. die Open Access Transformation und die Durchsetzung der FAIR Data Prinzipien mit dem Ziel der Nutzbarkeit. Hierbei müssen Data Stewards als Boundary Spanner ein Verständnis für alle Facetten der jetzigen und angestrebten Praktiken haben, was mögliche Unsicherheiten und Risiken in allen Akteursbereichen miteinschließt. Data Stewards müssen in der Lage sein, soziale und politische Dimensionen von aktuellen und zukünftigen Tätigkeitsspektren zu analysieren, diese Ergebnisse entsprechend in strategische Planungen umzusetzen und zwischen den Akteuren vermittelnd zu wirken. Weichen diese Dimensionen so weit voneinander ab, dass es zu pragmatischen Grenzen kommt, sollten Data Stewards als Boundary Spanner fähig sein, beispielsweise mit Hilfe von Plänen, Leitlinien oder Strategiepapieren Transformationsprozesse aufeinander abzustimmen und Zielvorstellungen zu vereinbaren.

Nachdem in These 1 hergeleitet wurde, dass das *Konzept* Data Stewardship unterschiedliche Akteure kollaborativ involvieren muss und die Überbrückung von unterschiedlich gearteten Boundaries bedarf, lässt sich somit die *Rolle* der Data Stewards als „Agents“ des Data Stewardships als Boundary Spanner beschreiben. Dies führt zu folgender These:

These 2. Data Stewards übernehmen die Rolle von Boundary Spannern, insbesondere im Bezug auf Übersetzungs- und Koordinationstätigkeiten. Sie sorgen dabei dafür, dass Kollaborationen zwischen unterschiedlichen Akteursgruppen insbesondere im Bezug auf Planung und Management, aber auch im Hinblick auf eine Veränderung von wissenschaftlichen Praktiken hin zu Open Science erfolgreich sind.

Als Boundary Spanner partizipieren Data Stewards somit insbesondere am *sharing* von Informationen (Hsiao, Tsai und Lee, 2012) und nehmen auch die Rolle des „Brokers“ (Pawlowski und Robey, 2004) ein. Der Dreiteilung der Aufgabenbereiche von Data Stewards in „embedded and operational“, „policy, strategy and coordination“ und „generic and advice“ aus dem Report von Verheul et al. (2019) kann im Hinblick auf Managementaufgaben („generic and advice“) und strategischen Planungs- und Veränderungsmaßnahmen („policy, strategy and coordination“) zugestimmt werden. Data Stewards an syntaktischen Grenzen sind dabei eher im Bereich „embedded and operational“ tätig.

7.3 Mögliche Boundary Objects

Die Verwendung von Boundary Objects bietet eine gute Möglichkeit, Boundaries zu überwinden und Kollaboration zu fördern. Dabei sollten die BOs so gestaltet sein, dass sich alle Akteure gleichermaßen darauf beziehen können. Dies kann jedoch nur funktionieren, wenn diese Objekte sowohl die Referenzbeziehungen auf das Objekt innerhalb der Sozialen Welt (beispielsweise im Bezug auf fachspezifisches Wissen) als beibehalten können auch als gemeinsames Bezugsobjekt aller beteiligten Akteure verwendet werden können (Meister, 2011, S.9). BOs können jedoch nicht unbedingt strategisch als solche konzipiert werden. Sie entstehen dadurch, dass sie in diverse Praktiken aller beteiligten Gruppen involviert werden. Nicht alle als BOs gewählten Artefakte werden daher auch als solche verwendet (Levina und Vaast, 2005). Zudem entstehen BOs innerhalb von Interaktionen häufig eher organisch, die Beschreibung als solche wird erst rückblickend und nicht durch die beteiligten Akteure vorgenommen (Hawkins, Pye und Correia, 2017). Semantische und pragmatische Boundaries können insbesondere von BOs profitieren, da durch sie eine „Auslagerung“ von (implizitem) Wissen in Form von Dekontextualisierung und Depersonalisierung erreicht werden kann. Die in Thesen 1 und 2 entwickelten Ansichten, dass Data Stewardship Tätigkeiten als Kollaborationen über Boundaries hinweg angesehen werden können, bei der Data Stewards als Boundary Spanner wirken, führt zu der Überlegung, welche Artefakte dabei als BOs eingesetzt werden. Artefakt wie kontrollierte Vokabulare, Metadatenschemata oder Ontologien werden insbesondere für Übertragungsfunktionen zwischen Infrastruktureinrichtungen und Forschenden eingesetzt. Jedoch sollten BOs in allen Akteursgruppen Verwendung finden. Demzufolge werden folgende drei Artefakte als BOs vorgeschlagen: Datenmanagementpläne, die FAIR Data Prinzipien und Policy Dokumente für Open Access/Open Science.

DMPs können als Planungswerkzeug für das Datenmanagement von allen drei Akteursgruppen verwendet werden, besonders wenn sie als sogenannte „*machine-actionable*“ DMPs konzipiert sind (Miksa et al., 2019). Bereits traditionelle DMPs können als koordinierende

Dokumente für die Planung des Datenmanagements angesehen werden. Zwar werden sie in der Regel von Forschenden ausgefüllt, jedoch passiert dies häufig in Zusammenarbeit mit den Infrastruktureinrichtungen. DMPs können dabei an semantischen Grenzen als gemeinsamer Referenzpunkt für alle Beteiligten bestehen, da Informationen zum Datenmanagement in eine strukturierte und für alle verständliche Form gebracht werden müssen. Hierbei entstehen jedoch meist statische Dokumente, die zwar in der Planung koordinationsfördernd wirken können, dann jedoch nicht weiter benutzt werden. Wenn DMPs maschinell ausführbar realisiert sind und als sogenannte „*living documents*“ im Verlauf von Forschungsprojekten ständig ergänzt werden, können sie als eine Art „Drehscheibenobjekt“ verstanden werden, welches von allen Akteuren fortlaufend benutzt werden kann und gleichzeitig eine Brückenfunktion anbietet.

Data Stewards können DMPs als BOs verwenden, um an semantischen Grenzen Anforderungen von Seiten der Forschungsförderer und Entscheidungsträger, Bedarfe und Planung von Seiten der Forschenden und Service-Angebote von Seiten der Infrastruktureinrichtungen zusammenzubringen. DMPs können dabei als *Standardized Forms* verwendet werden.

Die FAIR Data Prinzipien sind so formuliert, dass sie nicht als Standard fungieren sondern beschreiben wie Daten optimal maschinell auffindbar, zugänglich, interoperabel und nachnutzbar vorliegen sollen. Die bewusst offen gehaltene Formulierung der einzelnen Prinzipien führt dementsprechend dazu, dass die FAIR Data Principles in jeder Fachdisziplin entsprechend der dort vorliegenden Datentypen und Praktiken interpretiert werden können. Zum Beispiel ist Grundvoraussetzung für die Befolgung der Prinzipien eine Beschreibung von Daten mit adäquaten Metadaten. Welche Metadatenschemata verwendet werden sollten ist abhängig von der jeweiligen Fachdisziplin. Die FAIR Data Principles richten sich vor allem auf eine maschinell ausführbare Form des Umgang mit Daten aus. Dabei können die Prinzipien als BOs in Form von *Ideal Types* verwendet werden. Sie bilden, in für Menschen verständlicher Sprache und in abstrakter Form, das Ziel FAIR-konformer Daten ab und können als Referenzobjekte für unterschiedliche Bestrebungen auf dieses gemeinsame Ziel hin verwendet werden. Damit können die FAIR Data Principles sowohl an semantischen als auch an pragmatischen Grenzen wirken.

Data Stewards können diesen gemeinsamen Referenzpunkt benutzen, um einerseits Anforderungen an FAIR-konforme Daten zwischen den unterschiedlichen Akteuren zu übersetzen und andererseits um gemeinsame Prozesse hin einer „FAIRifizierung“ Daten zu orchestrieren. Die weit verbreitete Verwendung der FAIR Data Principles in der wissenschaftlichen Literatur seit ihrem Erscheinen²² ohne dass unbedingter Konsens darüber bestehen muss, was genau die einzelnen Prinzipien meinen, kann ein Hinweis auf die tatsächliche Verwendung der FAIR Data Principles als BO sein.

Unter Einbeziehung der Thesen 1 und 2 führt die Betrachtung von möglichen Boundary Objects und ihrem Einsatz im Data Stewardship zu folgender, dritten These:

In diesem Anhang wird die Verortung der im EOSCpilot entwickelten Data Stewardship Tätigkeiten analog zu ihrer Beschreibung in Kapitel 6 schematisch dargestellt. Die Kreise

²²Der Artikel (Wilkinson et al., 2016) zählt im Juli 2019 670 Zitation im Web of Science, 874 Zitationen in Crossref und hat eine Altmetrics-Score von 1343. Aktuell siehe: <https://www.nature.com/articles/sdata201618/metrics>

symbolisieren dabei die Sozialen Welten der Akteure (Infrastruktureinrichtungen = rot, Entscheidungsträger = grün, Forschende = blau). Der gelbe Kreis der unterschiedlichen Akteursgruppen gut umgesetzt werden können. Policies können dabei einen konzeptionellen und strategischen Rahmen abstecken (z.B. „freier Zugang zu wissenschaftlichen Publikationen“ in OA-Policies), der Praktiken über unterschiedliche Akteursgruppen hinaus legitimiert. Sie beinhalten dann einen symbolischen Wert für alle Akteursgruppen, der unter Umständen emotional aufgeladen ist. Briers und Chua (2001) nennen diese Art von BOs *visionary objects*. Policies können, wenn sie entsprechend gestaltet sind, als solche verwendet werden. Beispiele hierfür können die „Leitlinien zur Sicherung der Guten Wissenschaftlichen Praxis“ der DFG²³ aber auch institutionelle Open Access-Leitlinien sein.

Data Stewards können diese BOs insbesondere an pragmatischen Grenzen verwenden, um gemeinsame Werte und Ziele für unterschiedlicher Akteursgruppen zu vereinbaren. Zudem können Data Stewards, wenn sie an der Erstellung von Leitlinien beteiligt sind, einen aktiven Anteil am Design der Leitlinien leisten, mit dem Vorhaben sie als BOs einzusetzen.

Bei der Betrachtung von Personen und ihren Wirkungen an Grenzen stellt sich die Frage, ob Boundary Spanner nicht auch gleichzeitig als „menschliche“ Boundary Objects fungieren. Ursprünglich wird die Möglichkeit, dass auch Menschen BOs sein können von Star und Griesemer zurückgewiesen. Dies liegt daran, dass BOs passive Funktionen erfüllen müssen - daher auch der Name „Objekte“ - während Menschen als konsequent aktive Akteure angesehen werden (Star und Griesemer, 1989). In der Weiterentwicklung des Konzepts von BOs gibt es jedoch Hinweise darauf, dass „menschliche BOs“ existieren könnten. Insbesondere sogenannte „*versatile experts*“, also Personen, die mit ihrem Expertenwissen in unterschiedlichen Kontexten eingesetzt werden können, werden als BOs vorgeschlagen (Zdunczyk, 2006). Jedoch muss hierbei klar unterschieden werden zwischen Menschen als Grenzarbeiter und Grenzobjekten. Grenzarbeit verwendet Grenzobjekte, während Grenzobjekte selber keine eigenständige Grenzarbeit leisten können. Daher wurde der Begriff *Boundary Subject* vorgeschlagen. Boundary Subjects sind „aktive BOs“, die zwar ähnliche Charakteristika wie BOs aufweisen (unterschiedliche Bedeutungen in unterschiedlichen Welten mit einer gemeinsamen Kernbedeutung, Mediations- und Übersetzungsfunktion), sich jedoch aktiv in Kollaborationsprozesse einbringen und zwischen ihnen vermitteln. Boundary Subjects führen somit Boundary-Spanning-Aktivitäten aus, indem sie sich selber als BO verwenden. Boundary Subjects agieren an Grenzen und schaffen Verbindungen mit Hilfe von Kommunikation. Sie sind damit besonders gut geeignet um pragmatische Grenzen zu überwinden (Huzzard, Ahlberg und Ekman, 2010; Laine et al., 2016).

Unter Einbeziehung der Thesen 1 und 2 führt die Betrachtung von möglichen Boundary Objects und ihrem Einsatz im Data Stewardship zu folgender, dritten These:

These 3. Boundary Objects können insbesondere an syntaktischen und pragmatischen Grenzen für ein besseres Verständnis der Akteure und ihren jeweiligen sozialen Welten sowie Kollaborationsmaßnahmen eingesetzt werden. Im Data Stewardship können DMPs, die FAIR Data Prinzipien und Policies als BOs fungieren.

²³Siehe: https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/. Abgerufen am 21. Oktober 2019.

8 Fazit und Ausblick

Zusammenfassung

In dieser Arbeit wurden Data Stewardship und die Rolle von Data Stewards vor dem Hintergrund der Kollaboration zwischen Akteuren aus unterschiedlichen Sozialen Welten betrachtet. Forschungsdatenmanagement wurde dabei als soziotechnisches System angesehen in dem unterschiedliche Akteure in Beziehung treten und zusammenwirken müssen. Ziel der vorliegenden Arbeit war die Erarbeitung eines Referenzmodells durch die Übertragung von Theorien zu Kollaborationen in soziotechnischen Systemen und des Agierens an Boundaries an Grenzbereichen zwischen unterschiedlichen Sozialen Welten (*Boundaries*).

Dafür wurde zunächst die Begriffs- und Bedeutungsentwicklung von Data Stewardship in der Literatur nachvollzogen, um ein inhaltliches Verständnis herzustellen. Das Begriffsspektrum von Data Stewardship hat sich in den letzten Jahren gewandelt. Dies geschah besonders im Hinblick auf Nachhaltigkeit und Nachnutzbarkeit von FAIR-konform vorliegenden Daten. Der Begriff bezeichnete zu Anfang eine reine Langzeitarchivierung von Daten und veränderte sich hin zu einem ganzheitlicherem Verständnis einer Form des übergreifenden Datenmanagements mit dem Ziel langfristig nachnutzbarer und FAIR-konformer Datenpublikationen. Jedoch sind die Tätigkeitsbeschreibungen in den Data Stewardship Definitionen weiterhin nicht immer eindeutig von Aktivitäten im Datenmanagement oder Data Curation abgrenzbar. Aus der Literatur ließen sich folgende drei begriffliche Ausprägungen entnehmen:

- Data Stewardship als übergreifendes Konzept und zielgerichtete Menge an Handlungen,
- Data Stewardship als operative Beschreibungen von tatsächlichen Tätigkeiten, die von unterschiedlichen Akteuren im Forschungsprozess ausgeführt werden und
- Data Steward als Rollenbeschreibung für verantwortliche Personen als „Agents“ des Data Stewardships.

An einem Großteil von Tätigkeiten im Data Stewardship sind unterschiedliche Akteure beteiligt. Als gemeinsames Ziel können hierfür im Sinne der FAIR-Prinzipien auffindbare, zugängliche, interoperable und vor allem nachnutzbar vorliegende Daten angesehen werden. In bestehenden Theorien zu Kollaborationsprozessen zwischen Akteuren unterschiedlicher Sozialer Welten wird davon ausgegangen, dass unterschiedliche Formen von *Boundary-Work* diese Prozesse optimieren können. Boundary-Work bezeichnet dabei Überbrückungs- und Austauschfunktionen deren Durchführung Schwellen zwischen verschiedenen Akteursgruppen, z.B. in der Kommunikation, überwinden können.

Der Setzung im *Practical Turn* der Soziologie folgend wurde davon ausgegangen, dass Wissen sich in Praktiken wiederfindet. Daher wurden in dieser Arbeit 59 konkrete Tätigkeiten im Data Stewardship, die vom EU-Projekt EOSCpilot identifiziert wurden, den Akteuren Infrastruktureinrichtungen, Entscheidungsträgern und Forschenden zugeordnet. Dabei wurde beschrieben, welche Akteure an den jeweiligen Tätigkeiten beteiligt sind und wo es zu Kollaborationen kommt. Die Analyse zeigte, dass an den meisten Tätigkeiten im Data Stewardship mindestens zwei der drei betrachteten Akteure beteiligt sind. Dabei müssen

besonders häufig die Infrastruktureinrichtungen mit den Forschenden zusammenarbeiten. Die Entscheidungsträger sind dabei die einzige Akteursgruppe, die nur in Kollaboration mit anderen Akteuren am Data Stewardship beteiligt ist. Dies zeigt, dass Data Stewardship-Aktivitäten bislang primär Forschenden und der Infrastruktureinrichtungen zugeordnet werden, obwohl der Einfluss der Entscheidungsträger nicht unterschätzt werden sollte.

Data Stewardship kann als Boundary-Work im Sinne von Grenzüberbrückungen (*Boundary-Spanning*) betrachtet werden. Boundary-Spanning-Aktivitäten werden von designierten Personen an kognitiven Grenzen (*Knowledge Boundaries*) zwischen unterschiedlichen Sozialen Welten durchgeführt. Wie in dieser Arbeit dargestellt müssen dabei unterschiedliche Arten von Boundaries mit zunehmender Komplexität überwunden werden. Es handelt sich um syntaktische, semantische und pragmatische Boundaries. An syntaktischen Grenzen muss zwischen den Welten eine Übertragung stattfinden (Transfer). Hierfür wird eine gemeinsame lexikalische Grundlage benötigt, z.B. in Form von kontrollierten Vokabularen. An semantischen Grenzen muss zudem ein gemeinsames Verständnis von Informationen und Wissen durch Übersetzungsleistungen (Translate) hergestellt werden. Hilfreich können dabei Standardisierungsverfahren sein. Pragmatische Grenzen setzen eine Veränderung von Wissen, Praktiken und Zielen (Transform) voraus um eine Kollaboration zu ermöglichen. Dafür können konkrete Objekte, Modelle oder Strategien unterstützend verwendet werden. Boundary Spanner unterstützen diesen Prozess und wirken als Vermittler, Kommunikatoren und Mediatoren.

Bei der Betrachtung der Tätigkeiten konnten alle drei Arten von Grenzen identifiziert werden. Syntaktische Grenzen überwiegen im direkten Umgang mit Daten, beispielsweise in der Übertragung von Daten zwischen Forschenden und Infrastruktureinrichtungen. Semantische Boundaries tauchen insbesondere in Management- und Planungsprozessen auf. Pragmatische Boundaries finden sich in Transformationsprozessen mit dem Ziel von Open Science. Data Stewards übernehmen dabei unterstützende, koordinierende und vermittelnde Funktionen als Boundary Spanner. Hohes Verhandlungsgeschick, Empathie, ein tiefes Verständnis von bestehenden Arbeitskulturen und Vorgängen sowie Weitsicht im Sinne von Nachhaltigkeitsbestreben sind Voraussetzung für Data Stewards in der Rolle als Boundary Spanner.

Für die Arbeit an Boundaries können Boundary Objects eingesetzt werden, um als gemeinsamer Referenzpunkt zu dienen und so zwischen den Akteursgruppen zu vermitteln. Die hier vorgeschlagenen Artefakte (Datenmanagementpläne, Policies und die FAIR Data Prinzipien) zeigen durch ihre für die einzelnen Akteurgruppen flexible Interpretation sowie einem allgemeinen, sinnstiftenden Kern grundlegende Eigenschaften von BOs. Dabei können DMPs als *Standardized Forms*, Policies als *Visionary Objects* und die FAIR Data Principles als *Ideal Types* beschrieben werden. Sie können insbesondere an semantischen und pragmatischen Grenzen eingesetzt werden, um dort dekontextualisierend und depersonalisierend zu wirken und Übersetzungs- bzw. Transformationstätigkeiten zu vereinfachen.

Daher kann die Übertragbarkeit von bestehenden theoretischen Modellen zu Kollaborationen zwischen unterschiedlichen Sozialen Welten als durchführbar angesehen werden. In dieser Arbeit wurden dementsprechend drei Thesen entwickelt:

1. Data Stewardship findet an Boundaries statt.
2. Data Stewards agieren als Boundary Spanner.
3. Als Methode des Boundary Spannings können BOs verwendet werden. Beispiele hierfür sind DMPs, die FAIR Data Principles und Policies/Strategiepapiere.

Implikationen und Limitationen

Diese Arbeit schlägt eine neue Form der Betrachtung auf Forschungsdatenmanagement und die darunter fallenden Aufgaben vor einem soziotechnologischen Hintergrund vor. In Anbetracht bereits existierender Theorien zu Boundaries zwischen sozialen Welten und Boundary-Spanning-Aktivitäten zeigt sich hier eine in der Literatur bisher noch nicht dargestellte und damit völlig neue Perspektive auf den Umgang mit wissenschaftlichen Daten.

Diese Perspektive bietet die Möglichkeit, eine umfassendere Sicht auf das Datenmanagement zu entwickeln und so von bereits in anderen Kontexten, wie beispielsweise in der Organisationstheorie, erprobten Strategien und Ansätzen zu kollaborativem Arbeiten bei der Produktentwicklung sowie im Innovationsmanagement zu profitieren. Zudem kann die Perspektive dabei behilflich sein, epistemologische Grenzen im Forschungsdatenmanagement aufzuzeigen und Überwindungsstrategien zu entwickeln. Bei der Arbeit an Boundaries treten für alle Beteiligten auch Lerneffekte auf. Dabei handelt es sich um Identifikationsprozesse von unterschiedlichen Praktiken und ihren Relationen zueinander, um Koordinationsmechanismen und ihre Abstimmung aufeinander, um Perspektivenerweiterung in den jeweiligen Akteursgruppen durch die Reflektion der eigenen Praxis in Abgrenzung zu anderen Praktiken und mögliche Transformationsprozesse zur kollaborativen Schaffung neuer Praktiken (Akkerman und Bakker, 2011). Wenn diese Lernprozesse systematisch vor dem theoretischen Framework von Boundaries untersucht werden, besteht die Möglichkeit, Kompetenzfelder und auch „weiche“ Kriterien des Data Stewardships genauer zu untersuchen und auszudifferenzieren.

Zudem kann mit der neuen Perspektive auf das Datenmanagement die (neue) Berufsgruppe der Data Stewards genauer beschrieben werden. Der RfII führte kurz vor Beendigung dieser Arbeit bereits aus, dass Infrastruktureinrichtungen wie Bibliotheken und Rechenzentren in ihrer Struktur einem Veränderungsprozess unterliegen: *„Beschäftigte werden dort zu ‚Brokern‘ für die Vernetzung von Nutzerinnen und Nutzern sowie Nutzergruppen, für die digitale Prozessierung von Wissen und für die Nutzung von Datendiensten.“* (Rat für Informationsinfrastrukturen, 2019, S.8). Die Beschreibung von Personen, die als ‚Broker‘ tätig sind, kann in einem direkten Zusammenhang mit dem theoretischen Konzept von Data Stewards als Boundary Spanner gesehen werden. Die im RfII-Papier beschriebenen Aufgabenbereiche von Personen der Gruppe [B], die dort wirken, wo Pfadentscheidungen im Forschungsdatenlebenszyklus getroffen werden müssen, werden folgendermaßen beschrieben: *„Aufgaben, die einer forschungs- und fachnahen Expertise bedürfen, die den Forschungsprozess kollaborativ begleitet und/oder (etwa informationsinfrastrukturell) mitträgt; hier sind oftmals Personen mit Hochschulabschluss tätig.“* (ebd., S.7). Der RfII bezieht sich im Zuge der Entwicklung neuer Berufsfelder auf die Bezeichnungen „Data Curator/Librarian“ und „Di-

gital/Computational X”²⁴. Obwohl die Bezeichnung Data Steward oder Data Stewardship im gesamten Bericht nicht erwähnt wird, können die Ergebnisse dieser Arbeit dazu dienen, eine Diskussion anzustoßen, ob nicht ebendiese genau die Bedarfe erfüllen können, die aufgrund von fehlenden Berufsbildern, Kenntnissen oder Spezialisierungen bislang auftreten.

Nicht zuletzt kann durch den, zum Teil gesteuerten, Einsatz von unterschiedlichen Boundary Objects, eine koordinierende Arbeit zwischen allen Beteiligten maßgeblich gestärkt werden. Hierzu kann die theoretische Betrachtung unterschiedlicher Artefakte, wie zum Beispiel wie hier vorgeschlagen DMPs, die FAIR Data Principles und Policies, Auswirkung auf ihren praktischen Einsatz haben. Die strukturellen Anforderungen an BOs im Sinne von funktionierenden Übersetzungswerkzeugen können als Hintergrund dienen, um neue Services oder Strategien, wie zum Beispiel DMP-Tools oder Policies, zu entwickeln. Gleichzeitig kann der bewusste Einsatz von BOs hilfreich sein, auch implizite Praktiken, beispielsweise im Hinblick auf FAIR Data, zu untersuchen.

Jedoch zeigen sich in dieser Arbeit auch eine Reihe von Limitationen. Erstens sind, wie in Kapitel 4 beschrieben, weder die Definitionen des Konzepts Data Stewardship noch des Berufsbildes bzw. der Rolle der Data Stewards vollständig ausgereift. Somit beziehen sich die in dieser Arbeit untersuchten Tätigkeiten zu einem überwiegenden Teil auf theoriebasierte Ergebnisse des EOSCpilot Projekts. Wie genau sich diese in die Praxis übertragen lassen, muss sich in der Zukunft noch zeigen. Im Abschlussbericht des Projektes²⁵ wurden aus den 59 beschriebenen Tätigkeiten zehn Schlüsselkompetenzen sowie eine Reihe von „*skill profiles*” erstellt (Whyte et al., 2019). Inwiefern diese sich vor dem hier betrachteten Hintergrund von Boundary-Work einordnen lassen, muss noch untersucht werden. Die vorgenommenen Analyse in dieser Arbeit dient dabei zur Theoriebildung für nachfolgende Forschungsansätze. Keines der Ergebnisse ist empirisch belegt. Zweitens ist sicherlich nicht zu unterschätzen, dass Untersuchungen zu Boundaries und insbesondere Boundary-Spanning-Aktivitäten zu einem großen Anteil auf Beobachtungen im Kontext von Unternehmen beruhen. Wie weit eine Übertragbarkeit auf den Wissenschaftsbetrieb, wie es in dieser Arbeit vorgeschlagen wurde, möglich ist kann sich erst durch praktische Anwendung des hier entwickelten theoretischen Konzepts erweisen. Zusätzlich kann diskutiert werden, ob das Konzept der Boundary Objects, obwohl es weitreichend verwendet wird, als eine abschließend hinreichende Perspektive angesehen werden kann. Die Verwendung von Boundary Objects ist *eine* Form der Betrachtung von koordinierenden Maßnahmen zur Kollaboration. Andere Betrachtungsmöglichkeiten, wie z.B. „*epistemic things*” (Rheinberger, 1997) oder andersgeartete Ansätze können zu anderen Ergebnissen führen. Außerdem sollte, wenn Data Stewardship Aktivitäten eingebettet in ein soziotechnologisches System gesehen werden, die technologische Seite des Datenmanagements genauer betrachtet werden. Welchen Einfluss hat die Qualität der verwendeten technischen Systeme auf die Überbrückung von Knowledge Boundaries? Fragestellungen wie diese könnten in weiteren Forschungsarbeiten bearbeitet werden.

Diese Arbeit hat zeigt, dass das Data Stewardship über eine rein infrastrukturelle oder forschungsgetriebene Perspektive hinaus als Zusammenspiel unterschiedlicher Akteure und

²⁴X steht dabei für eine bestimmte Fachdisziplin. Beispiel: Digital Humanist, Computational Biologist.

²⁵Der Bericht erschien nachdem die Analyse dieser Arbeit bereits auf Grundlage der vorherigen Ergebnisse abgeschlossen war

eingebettet in soziale Geflechte als Fallbeispiel für soziologische Untersuchungen des komplexen Betriebs „Wissenschaft“ als interessantes Beispiel dient.

Ausblick

Um das in dieser Arbeit erstellte theoretische Modell und die daraus abgeleiteten Thesen empirisch zu überprüfen müssten im nächsten Schritt konkrete Untersuchungen „im Feld“ durchgeführt werden. Hierfür bieten sich Methoden aus der Soziologie, wie die (teilnehmende) Beobachtung über einen längeren Zeitraum an. Diese könnten sich sowohl auf konkrete Abläufe des Forschungsdatenmanagements in einer Einrichtung beziehen, um dort Boundaries zu verorten, als auch, idealerweise, an bereits existierenden Data Stewardship Programmen angesiedelt werden. Auf Grundlage dieser Methode könnten auch bereits verwendete Boundary Objects, die in dieser Arbeit nicht betrachtet werden konnten, erkannt werden. Zudem ließen sich konkrete Tätigkeiten, Relationsgeflechte und unbekannte Kollaborationspotentiale herausarbeiten. Die Betrachtung des „Ökosystems Datenmanagement“ auf Grundlage eines Social-Worlds-Frameworks fordert dabei eine interdisziplinäre Perspektive. Besonders fruchtbar könnte hierbei die Zusammenarbeit zwischen Informationswissenschaftler*innen, Wissenschaftssoziolog*innen und Forschenden mit dem Schwerpunkt „Digital Sociology“ sein.

Zu überlegen wäre auch, ob die betrachteten Akteure anhand ihrer Statusgruppen betrachtet werden müssten. Es wäre potentiell aufschlussreicher, sogenannte „Communities of data“ (in Anlehnung an den Term „Communities of practice“) zu betrachten. Hierbei stünden Daten als gemeinsame Bezugsobjekte stärker im Vordergrund als institutionell lokalisierte Abläufe des Datenmanagements, wie sie in Data Life Cycle Modellen beschrieben sind. Diese Communities ließen sich beispielsweise über Co-Citation Analysen von Datensätzen ausfindig machen. Da sich die Zitation von Daten bislang allerdings in den meisten Wissenschaftsdiziplinen bei weitem noch nicht durchgesetzt hat liegt dieser Ansatz bisher allerdings noch in der Zukunft.

Weitere Möglichkeiten liegen in der Betrachtung von Lerneffekten an Boundaries und dem Bezug zu „Data Literacy“ sowie wissenskulturell orientierte Informationspraktika. Wenn diese Lerneffekte sowohl für die Akteursgruppen als auch für Data Stewards in weiteren Forschungsarbeiten systematisch untersucht würden, könnte man die Ausbildung im Bereich des Datenmanagements entsprechend ausrichten. Hierzu zählen sowohl praktische Fähigkeiten als auch „Soft Skills“ wie interpersonelle Kompetenzen.

Eine andere Forschungsfrage, die sich im Bezug auf Data Stewardship als Boundary Work ergibt ist die Frage nach Vertrauen: Welche Rolle spielt Vertrauen zwischen Data Stewards und allen anderen Akteursgruppen? Wie lässt sich dieses Vertrauen herstellen und beibehalten? Welche Rolle spielen vertrauenswürdige Technologien dabei? Können Boundary Objects zur Vertrauensbildung eingesetzt werden und welche könnten das sein? Wie werden soziale Infrastrukturen hierdurch beeinflusst?

Insgesamt lässt sich festhalten, dass Data Stewards und das Data Stewardship in Zukunft eine

wichtige Rolle im Hinblick auf eine nachhaltige, transparente und nachnutzbare Wissenschaft im Sinne der FAIR Data Principles spielen werden. Forschung zum Thema interdisziplinäre Kollaborationen, Kompetenzausbildung und Koordinierungsprozessen kann in den nächsten Jahren zu dieser Entwicklung maßgeblich beitragen.

Literatur

- Abele-Brehm, Andrea E. et al. (2019). „Attitudes Toward Open Science and Public Data Sharing“. In: *Social Psychology* 50 (4), S. 1–9. DOI: <https://doi.org/10.1027/1864-9335/a000384>.
- Addison, Aaron; Moore, Jennifer; Hudson-Vitale, Cynthia (2015). „Forging Partnerships: Foundations of Geospatial Data Stewardship“. In: *Journal of Map & Geography Libraries* 11 (3), S. 359–375. DOI: <https://doi.org/10.1080/15420353.2015.1054544>.
- AG Open Science (2019). *Open Science*. URL: <https://www.ag-openscience.de/openscience/> (Abgerufen am 25. 05. 2019).
- Akkerman, Sanne F.; Bakker, Arthur (2011). „Boundary Crossing and Boundary Objects“. In: *Review of Educational Research* 81 (2), S. 132–169. DOI: <https://doi.org/10.3102/0034654311404435>.
- Association of Research Libraries (2006). *ARL Bimonthly Report*. URL: <http://old.arl.org/resources/pubs/br/br249.shtml> (Abgerufen am 30. 03. 2019).
- Baker, Karen S.; Chandler, Cynthia L. (2008). „Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics“. In: *Deep Sea Research Part II: Topical Studies in Oceanography* 55 (18), S. 2132–2142. DOI: <https://doi.org/10.1016/j.dsr2.2008.05.009>.
- Baker, Karen S.; Yarmey, Lynn (2009). „Data Stewardship: Environmental Data Curation and a Web-of-Repositories“. In: *International Journal of Digital Curation* 4 (2), S. 12–27. DOI: <https://doi.org/10.2218/ijdc.v4i2.90>.
- Bechky, Beth A. (2003). „Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor“. In: *Organization Science* 14 (3), 312–330. DOI: <https://doi.org/10.1287/orsc.14.3.312.15162>.
- Benn, Suzanne; Edwards, Melissa; Angus-Leppan, Tamsin (2013). „Organizational Learning and the Sustainability Community of Practice“. In: *Organization Environment* 26 (2), S. 184–202. DOI: <https://doi.org/10.1177/1086026613489559>.
- Berle, Adolf A.; Means, Gardiner C. (1991). *The modern corporation and private property*. New Brunswick, N.J., U.S.A: Transaction Publishers.
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-term Access to Digital Information*. La Jolla, Calif.: Blue Ribbon Task Force on Sustainable Digital Preservation und Access. URL: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (Abgerufen am 01. 04. 2019).
- Boeckhout, Martin; Zielhuis, Gerhard A.; Bredenoord, Annelien L. (2018). „The FAIR guiding principles for data stewardship: fair enough?“ In: *European Journal of Human Genetics* 26 (7), S. 931–936. DOI: <https://doi.org/10.1038/s41431-018-0160-0>.
- Bongaerts, Gregor (2007). „Soziale Praxis und Verhalten – Überlegungen zum Practice Turn in Social Theory“. In: *Zeitschrift für Soziologie* 36 (4). DOI: <https://doi.org/10.1515/zfsoz-2007-0401>.
- Borgman, Christine L. (2012). „The Conundrum of Sharing Research Data“. In: *Journal of the American Society for Information Science and Technology* 63 (6), S. 1059–1078. URL: <https://papers.ssrn.com/abstract=2073876> (Abgerufen am 15. 04. 2019).

- Borgman, Christine L (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA., U.S.A.: The MIT Press.
- Bowker, Geoffrey C.; Star, Susan Leigh (1999). *Sorting things out: classification and its consequences*. Cambridge, MA., U.S.A.: The MIT Press.
- Bowker, Geoffrey C. et al. (2009). „Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment“. In: *International Handbook of Internet Research*. Hrsg. von Jeremy Hunsinger; Lisbeth Klastrup; Mathew Allen. Springer Netherlands, 97–117. DOI: https://doi.org/10.1007/978-1-4020-9789-8_5.
- Briers, Michael; Chua, Wai Fong (2001). „The role of actor-networks and boundary objects in management accounting change: a field study of an implementation of activity-based costing“. In: *Accounting, organizations and society* 26 (3), S. 237–269. DOI: [https://doi.org/10.1016/S0361-3682\(00\)00029-5](https://doi.org/10.1016/S0361-3682(00)00029-5).
- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (2011). *Handbuch Forschungsdatenmanagement*. Bock+ Herchen.
- Carlile, Paul R. (2002). „A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development“. In: *Organization Science* 13 (4), S. 442–455. DOI: <https://doi.org/10.1287/orsc.13.4.442.2953>.
- (2004). „Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries“. In: *Organization Science* 15 (5), S. 555–568. DOI: <https://doi.org/10.1287/orsc.1040.0094>.
- Clarke, Adele E; Star, Susan Leigh (2008). „The social worlds framework: A theory/methods package“. In: *The handbook of science and technology studies*. Hrsg. von Ulrike Felt; Rayvon Fouche; Clark A. Miller. 3. Aufl. The MIT Press, S. 113–137.
- Collins, Sandra et al. (2018). *Turning FAIR data into reality: final report from the European Commission Expert Group on FAIR data*. URL: <https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283> (Abgerufen am 08.09.2018).
- Dawes, Sharon S. (1996). „Interagency information sharing: Expected benefits, manageable risks“. In: *Journal of Policy Analysis and Management* 15 (3), S. 377–394. DOI: [https://doi.org/10.1002/\(SICI\)1520-6688\(199622\)15:3<377::AID-PAM3>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1520-6688(199622)15:3<377::AID-PAM3>3.0.CO;2-F).
- Department of Defense (1998). *Data Standardization Procedures, DoD 8320.1-M-1*. URL: <https://apps.dtic.mil/docs/citations/ADA343735> (Abgerufen am 05.04.2019).
- Dolzycka, Dominika et al. (März 2019). *Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement*. DOI: <https://doi.org/10.5281/zenodo.2581292>.
- Donaldson, Lex; Davis, James H. (1991). „Stewardship Theory or Agency Theory: CEO Governance and Shareholder Returns“. In: *Australian Journal of Management* 16 (1), S. 49–64. DOI: <https://doi.org/10.1177/031289629101600103>.
- Dravis, Frank (2004). „Data quality strategy: a step-by-step approach“. In: *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)* (Cambridge, Massachusetts, USA, 7. Nov. 2004–7. Nov. 2004), S. 27–43. URL: <http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202004/Papers/DQStrategy.pdf> (Abgerufen am 03.04.2019).
- Duerr, Ruth et al. (2004). „Challenges in long-term data stewardship“. In: *1st IEEE Conference on Mass Storage Systems and Technologies / 12th NASA Goddard Conference on Mass Storage Systems and Technologies* (Greenbelt, Maryland, USA, 13. Apr.

- 2004–16. Apr. 2004), S. 101–121. URL: <https://pdfs.semanticscholar.org/f7e8/be9df6138f34e6cc3a0d02bc68107ac0192f.pdf> (Abgerufen am 30.03.2019).
- European Commission (2016). *Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud*. Luxembourg. URL: <http://bookshop.europa.eu/uri?target=EUB:NOTICE:KI0116872:EN:HTML> (Abgerufen am 23.08.2018).
- (2018a). *Commission Staff Working Document: Implementation Roadmap for the European Open Science Cloud*. URL: https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf (Abgerufen am 18.04.2019).
- (2018b). *Cost-benefit analysis for FAIR research data : cost of not having FAIR research data*. Luxembourg. URL: <https://publications.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en> (Abgerufen am 18.04.2019).
- (2018c). *Prompting an EOSC in practice : final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), 2018*. URL: <https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1> (Abgerufen am 17.04.2019).
- Faraj, Samer; Yan, Aimin (2009). „Boundary work in knowledge teams.“ In: *Journal of Applied Psychology* 94 (3), S. 604–617. DOI: <https://doi.org/10.1037/a0014367>.
- Fecher, Benedikt et al. (2017). „A reputation economy: how individual reward considerations trump systemic arguments for open access to data“. In: *Palgrave Communications* 3 (17051), S. 1–10. DOI: <https://doi.org/10.1057/palcomms.2017.51>.
- Fleming, Lee; Waguespack, David M. (2007). „Brokerage, Boundary Spanning, and Leadership in Open Innovation Communities“. In: *Organization Science* 18 (2), S. 165–180. DOI: <https://doi.org/10.1287/orsc.1060.0242>.
- Friedlander, Amy; Adler, Prudence (2006). *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering. A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships–The Role of Academic Libraries in the Digital Data Universe (Arlington, Virginia, September 26-27, 2006)*. Association of Research Libraries. URL: <https://eric.ed.gov/?id=ED528649> (Abgerufen am 02.04.2019).
- Fujimura, Joan H. (1992). „Crafting Science: Standardized Packages, Boundary Objects, and Translation.“ In: *Science as Practice and Culture*. Hrsg. von Andrew Pickering. University of Chicago Press, S. 168–211.
- Gal, Uri; Yoo, Youngjin; Boland, Richard J. (2005). „The dynamics of boundary objects, social infrastructures and social identities“. In: *Sprouts: Working Papers on Information Systems* 4 (11), S. 194–206. URL: https://aisel.aisnet.org/sprouts_all/67/ (Abgerufen am 10.04.2019).
- Garrety, Karin; Badham, R. (2000). „The Politics of Sociotechnical Intervention: An Interactionist View“. In: *Faculty of Commerce - Papers (Archive)*. URL: <https://ro.uow.edu.au/commpapers/156> (Abgerufen am 10.04.2019).
- Gieryn, Thomas F (1983). „Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists“. In: *American sociological review* 48 (6), S. 781–795. DOI: <https://doi.org/10.2307/2095325>.

- Gregory, Kathleen M et al. (2019). „Understanding data search as a socio-technical practice“. In: *Journal of Information Science*, S. 1–17. DOI: <https://doi.org/10.1177/0165551519837182>.
- Harris, Frances; Lyon, Fergus (2013). „Transdisciplinary environmental research: Building trust across professional cultures“. In: *Environmental Science & Policy* 31, 109–119. DOI: <https://doi.org/10.1016/j.envsci.2013.02.006>.
- Hartmann, Niklas K.; Jacob, Boris; Weiß, Nadin (2019). *RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement*. DOI: <https://doi.org/10.5281/zenodo.2549344>.
- Hawkins, Beverley; Pye, Annie; Correia, Fernando (2017). „Boundary objects, power, and learning: The matter of developing sustainable practice in organizations“. In: *Management Learning* 48 (3), S. 292–310. DOI: <https://doi.org/10.1177/1350507616677199>.
- Henty, Margaret (2008). *Dreaming of Data: The Library's Role in Supporting e-Research and Data Management*. URL: <http://hdl.handle.net/1885/47617>.
- Higgins, Sarah (2011). „Digital Curation: The Emergence of a New Discipline“. In: *International Journal of Digital Curation* 6 (2), S. 78–88.
- Hodson, Simon et al. (2018). *FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data*. DOI: <https://doi.org/10.5281/zenodo.1285290>.
- Hsiao, Ruey-Lin; Tsai, Dun-Hou; Lee, Ching-Fang (2012). „Collaborative knowing: the adaptive nature of cross-boundary spanning“. In: *Journal of management studies* 49 (3), S. 463–491. DOI: <https://doi.org/10.1111/j.1467-6486.2011.01024.x>.
- Huvila, Isto et al. (2017). „Boundary objects in information science“. In: *Journal of the Association for Information Science and Technology* 68 (8), S. 1807–1822. DOI: <https://doi.org/10.1002/asi.23817>.
- Huzzard, Tony; Ahlberg, Beth Maina; Ekman, Marianne (2010). „Constructing interorganizational collaboration: The action researcher as boundary subject“. In: *Action Research* 8 (3), S. 293–314. DOI: <https://doi.org/10.1177/1476750309335206>.
- Hörster, Reinhard; Köngeter, Stefan; Müller, Burkhard (2013). „Grenzobjekte und ihre Erfahrbarkeit in sozialen Welten“. In: *Grenzobjekte: Soziale Welten und ihre Übergänge*. Hrsg. von Reinhard Hörster; Stefan Köngeter; Burkhard Müller. Wiesbaden: Springer Fachmedien, S. 11–36. DOI: https://doi.org/10.1007/978-3-531-18953-6_1.
- Karasti, Helena; Baker, Karen S.; Halkola, Eija (2006). „Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network“. In: *Computer Supported Cooperative Work (CSCW)* 15 (4), S. 321–358. DOI: <https://doi.org/10.1007/s10606-006-9023-2>.
- Kellogg, Katherine C.; Orlikowski, Wanda J.; Yates, JoAnne (2006). „Life in the Trading Zone: Structuring Coordination Across Boundaries in Postbureaucratic Organizations“. In: *Organization Science* 17 (1), S. 22–44. DOI: <https://doi.org/10.1287/orsc.1050.0157>.
- Kindling, Maxi; Schirmbacher, Peter; Simukovic, Elena (2013). „Forschungsdatenmanagement an Hochschulen: das Beispiel der Humboldt-Universität zu Berlin“. In: *LIBREAS. Library Ideas* (23). DOI: <https://doi.org/10.18452/9041>.
- Laine, Teemu et al. (2016). „Boundary subjects and boundary objects in accounting fact construction and communication“. In: *Qualitative Research in Accounting & Management* 13 (3), S. 303–329. DOI: <https://doi.org/10.1108/QRAM-09-2015-0085>.

- Lamont, Michèle; Molnár, Virág (2002). „The study of boundaries in the social sciences“. In: *Annual review of sociology* 28(1), S. 167–195. DOI: <https://doi.org/10.1146/annurev.soc.28.110601.141107>.
- Latour, Bruno (1996). „On actor-network theory: A few clarifications“. In: *Soziale Welt* 47(4), S. 369–381. URL: <https://www.jstor.org/stable/40878163> (Abgerufen am 04.07.2019).
- Levina, Natalia; Vaast, Emmanuelle (2005). „The emergence of boundary spanning competence in practice: implications for implementation and use of information systems“. In: *MIS quarterly* 29(2), S. 335–363. DOI: <https://doi.org/10.2307/25148682>.
- Li, Xin et al. (2011). „Toward an improved data stewardship and service for environmental and ecological science data in West China“. In: *International Journal of Digital Earth* 4(4), S. 347–359. DOI: <https://doi.org/10.1080/17538947.2011.558123>.
- Mariam-Webster.com (2019). *Stewardship*. URL: <https://www.merriam-webster.com/dictionary/stewardship> (Abgerufen am 23.07.2019).
- Meister, Martin (2011). „Soziale Koordination durch Boundary Objects“. Dissertation. Technische Universität Berlin. URL: <https://d-nb.info/1017839786/34>.
- Miksa, Tomasz et al. (März 2019). „Ten principles for machine-actionable data management plans“. In: *PLOS Computational Biology* 15(3), S. 1–15. DOI: <https://doi.org/10.1371/journal.pcbi.1006750>.
- Mons, Barend et al. (2017). „Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud“. In: *Information Services & Use* 37(1), S. 49–56. DOI: <https://doi.org/10.3233/ISU-170824>.
- Moore, Reagan (2008). „Towards a Theory of Digital Preservation“. In: *International Journal of Digital Curation* 3(1), S. 63–75.
- National Digital Stewardship Alliance (2014). *A report on the challenges, opportunities, gaps, emerging trends, and key areas for research and development that support the national capacity for digital stewardship*. URL: <http://www.digitalpreservation.gov/documents/2015NationalAgenda.pdf> (Abgerufen am 28.05.2019).
- National Research Council (2007). *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*. Washington, D.C.: National Academies Press. DOI: <https://doi.org/10.17226/12017>.
- Neuroth, Heike et al. (2018). „Aktives Forschungsdatenmanagement“. In: *ABI Technik* 38(1), S. 55–64. DOI: <https://doi.org/10.1515/abitech-2018-0008>.
- Neuroth, Heike et al. (2019). *Embedded data stewardship: A community-driven agile self-assessment framework for monitoring and improving the quality of research data management*. Poster auf der 14th International Digital Curation Conference (idcc). URL: http://www.dcc.ac.uk/sites/default/files/documents/IDCC19/211_Embedded%20Data%20Stewardship%20A%20community%20driven%20agile%20self-assessment%20framework%20for%20monitoring%20and%20improving%20the%20quality%20of%20research%20data%20management.pdf.
- Neylon, Cameron (2017). *Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers*. DOI: <https://doi.org/10.3897/rio.3.e21705>.
- NOAA (2005). *Scientific Data Stewardship (SDS) Implementation Plan*. URL: <ftp://ftp.ncdc.noaa.gov/pub/data/sds/SDS.implementation.plan-V3.1-2-18-05.pdf> (Abgerufen am 28.05.2019).

- Palmer, Carole et al. (2013). „Foundations of Data Curation: The Pedagogy and Practice of 'Purposeful Work' with Research Data“. In: *Archives Journal* 3. URL: <http://hdl.handle.net/2142/78099>.
- Pawlowski, Suzanne; Robey, Daniel (Dez. 2004). „Bridging User Organizations: Knowledge Brokering and the Work of Information Technology Professionals“. In: *Management Information Systems Quarterly* 28 (4). URL: <https://aisel.aisnet.org/misq/vol28/iss4/6>.
- Pels, Peter et al. (2018). „Data management in anthropology: the next phase in ethics governance?“ In: *Social Anthropology* 26 (3), S. 391–413. DOI: <https://doi.org/10.1111/1469-8676.12526>.
- Peng, Ge (2018). „The State of Assessing Data Stewardship Maturity – An Overview“. In: *Data Science Journal* 17 (7), S. 1–12. DOI: <https://doi.org/10.5334/dsj-2018-007>.
- Peng, Ge et al. (2015). „A Unified Framework for Measuring Stewardship Practices Applied to Digital Environmental Datasets“. In: *Data Science Journal* 13, S. 231–253. DOI: <https://doi.org/10.2481/dsj.14-049>.
- Peng, Ge et al. (Mai 2016). „Scientific Stewardship in the Open Data and Big Data Era : Roles and Responsibilities of Stewards and Other Major Product Stakeholders“. In: *D-Lib Magazine* 22 (5/6). DOI: <https://doi.org/10.1045/may2016-peng>.
- Peng, Ge et al. (2018). „A Conceptual Enterprise Framework for Managing Scientific Data Stewardship“. In: *Data Science Journal* 17 (15), S. 1–17. DOI: <https://doi.org/10.5334/dsj-2018-015>.
- Piowar, Heather A. (2011). „Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data“. In: *PLoS ONE* 6 (7), e18657. DOI: <https://doi.org/10.1371/journal.pone.0018657>.
- Plotkin, David (2014). *Data stewardship: an actionable guide to effective data management and data governance*. Amsterdam ; Boston: Elsevier/Morgan Kaufman.
- Polanyi, Michael; Nye, Mary Jo (2015). *Personal knowledge: towards a post-critical philosophy*. Enlarged edition. Chicago: University of Chicago Press.
- Polman, Joseph L.; Hope, Jennifer M.G. (2014). „Science news stories as boundary objects affecting engagement with science“. In: *Journal of Research in Science Teaching* 51 (3), S. 315–341. DOI: <https://doi.org/10.1002/tea.21144>.
- Rat für Informationsinfrastrukturen (2016). *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen.
- (2017a). *Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich. Bericht und Anregungen*. Diskussionspapier. URL: <http://www.rfii.de/?wpdmdl=2346> (Abgerufen am 17.06.2019).
- (2017b). *Schritt für Schritt – oder: Was bringt wer mit? Ein Diskussionsimpuls zu Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI)*. Diskussionspapier. URL: <http://www.rfii.de/?wpdmdl=2269> (Abgerufen am 17.06.2019).
- (2018a). *In der Breite und forschungsnah: Handlungsfähige Konsortien Dritter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland*. URL: www.rfii.de/?wpdmdl=3509 (Abgerufen am 17.06.2019).

-
- (2018b). *Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland*. URL: <http://www.rfii.de/?wpdmdl=2529> (Abgerufen am 17.06.2019).
 - (2019). *Digitale Kompetenzen - dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft*. Göttingen. URL: <http://www.rfii.de/?p=3883> (Abgerufen am 05.07.2019).
 - Reckwitz, Andreas (2003). „Grundelemente einer Theorie sozialer Praktiken“. In: *Zeitschrift für Soziologie* 32 (4), S. 282–301. DOI: <https://doi.org/10.1515/zfsoz-2003-0401>.
 - Reiger, Horst (2009). „Symbolischer Interaktionismus“. In: *Qualitative Marktforschung: Konzepte – Methoden – Analysen*. Hrsg. von Renate Buber; Hartmut H. Holzmüller. Wiesbaden: Gabler, S. 137–156. DOI: https://doi.org/10.1007/978-3-8349-9441-7_9.
 - Rezazade M., Mohammad H.; Hawkins, Matthew A. (2012). „Knowledge boundary spanning process: synthesizing four spanning mechanisms“. In: *Management Decision* 50 (10), S. 1800–1815. DOI: <https://doi.org/10.1108/00251741211279611>.
 - Rheinberger, Hans-Jörg (1997). „Toward a history of epistemic things : synthesizing proteins in the test tube“. In: URL: <http://hdl.handle.net/11858/00-001M-0000-002A-AAA5-F>.
 - Rodrigo, Allen et al. (Jan. 2013). „Science Incubators: Synthesis Centers and Their Role in the Research Ecosystem“. In: *PLOS Biology* 11 (1), e1001468. DOI: <https://doi.org/10.1371/journal.pbio.1001468>.
 - Rolando, Lizzy et al. (2013). *Institutional Readiness for Data Stewardship: Findings and Recommendations from the Research Data Assessment*. URL: <http://hdl.handle.net/1853/48188>.
 - Rosenbaum, Sara (2010). „Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access“. In: *Health Services Research* 45 (5 Pt 2), S. 1442–1455. DOI: <https://doi.org/10.1111/j.1475-6773.2010.01140.x>.
 - Schimank, Uwe (2010). *Handeln und Strukturen: Einführung in die akteurtheoretische Soziologie*. Beltz Juventa.
 - (2015). „Governance der Wissenschaft“. In: *Handbuch Wissenschaftspolitik*. Hrsg. von Dagmar Simon; Andreas Knie; Stefan Hornbostel. Wiesbaden: Springer Fachmedien Wiesbaden, S. 1–15. DOI: https://doi.org/10.1007/978-3-658-05677-3_4-1.
 - Scholtens, Salome et al. (2019a). *Life sciences data steward function matrix*. DOI: <https://doi.org/10.5281/zenodo.2561723>.
 - (2019b). *Towards a community-endorsed data steward profession description for life science research - Poster*. DOI: <https://doi.org/10.5281/zenodo.2616477>.
 - Sciences (US), National Academy of (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington (DC): National Academies Press. URL: <http://www.ncbi.nlm.nih.gov/books/NBK215264/> (Abgerufen am 05.12.2018).
 - Shalini Kurapati (2019). *Becoming a data steward*. URL: <https://blogs.lse.ac.uk/impactofsocialsciences/2019/04/30/becoming-a-data-steward/> (Abgerufen am 11.05.2019).
 - Staiger, Christine (2019). *FAIR data stewardship - Introduction*. DOI: <https://doi.org/10.5281/zenodo.2585691>.

- Star, Susan Leigh (2010). „This is Not a Boundary Object: Reflections on the Origin of a Concept“. In: *Science, Technology, & Human Values* 35(5), S. 601–617. DOI: <https://doi.org/10.1177/0162243910377624>.
- Star, Susan Leigh; Griesemer, James R (1989). „Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39“. In: *Social studies of science* 19(3), S. 387–420. DOI: <https://doi.org/10.1177/030631289019003001>.
- Star, Susan Leigh; Ruhleder, Karen (1996). „Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces“. In: *Information Systems Research* 7(1), S. 111–134. DOI: <https://doi.org/10.1287/isre.7.1.111>.
- Steinhof, Christoph (2018). „Erfolgskriterien von Forschungsdatenrepositorien und deren Relevanz für verschiedene Stakeholder-Gruppen“. Masterthesis. Fachhochschule Potsdam. URL: [urn:nbn:de:kobv:525-19521](https://nbn-resolving.org/urn:nbn:de:kobv:525-19521).
- Stichweh, Rudolf (1992). „The Sociology of Scientific Disciplines: On the Genesis and Stability of the Disciplinary Structure of Modern Science“. In: *Science in Context* 5(1), 3–15. DOI: <https://doi.org/10.1017/s0269889700001071>.
- Strauss, Anselm (1978). „A social world perspective“. In: *Studies in symbolic interaction* 1(1), S. 119–128.
- Tammaro, Anna Maria; Casarosa, Vittore (2018). „Who Is the Data Curator? Defining a Vocabulary“. In: *Digital Libraries and Multimedia Archives*. Hrsg. von Giuseppe Serra; Carlo Tasso. Bd. 806. Cham: Springer International Publishing, S. 249–255. DOI: https://doi.org/10.1007/978-3-319-73165-0_25.
- Tammaro, Anna Maria et al. (2017). „Understanding roles and responsibilities of data curators: an international perspective“. In: *Libellarium: journal for the research of writing, books, and cultural heritage institutions* 9(2). DOI: <https://doi.org/10.15291/libellarium.v9i2.286>.
- Tenopir, Carol et al. (2014). „Research data management services in academic research libraries and perceptions of librarians“. In: *Library & Information Science Research* 36(2), S. 84–90. DOI: <https://doi.org/10.1016/j.lisr.2013.11.003>.
- Teperek, Marta (2018). *Data Stewardship – towards cultural change in data management*. DOI: <https://doi.org/10.5281/zenodo.1485805>.
- Teperek, Marta; Dunning, Alastair (2019). *Data (and code) roles of the future at TU Delft*. DOI: <https://doi.org/10.5281/zenodo.3256576>.
- Teperek, Marta et al. (2018). „Data Stewardship addressing disciplinary data management needs“. In: *International Journal of Digital Curation* 13(1), S. 141–149. DOI: <https://doi.org/10.2218/ijdc.v13i1.604>.
- Trist, E. L.; Bamforth, K. W. (1951). „Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System“. In: *Human Relations* 4(1), S. 3–38. DOI: <https://doi.org/10.1177/001872675100400101>.
- Trompette, Pascale; Vinck, Dominique (2009). „Revisiting the notion of Boundary Object“. In: *Revue d'anthropologie des connaissances* 3(1), S. 3–25. URL: <https://www.cairn.info/revue-anthropologie-des-connaissances-2009-1-page-3.htm> (Abgerufen am 17. 04. 2019).

- Tushman, Michael L.; Scanlan, Thomas J. (1981). „Boundary Spanning Individuals: Their Role in Information Transfer and Their Antecedents“. In: *Academy of Management Journal* 24 (2), S. 289–305. DOI: <https://doi.org/10.5465/255842>.
- Velte, Patrick (2010). „Stewardship-Theorie“. In: *Zeitschrift für Planung & Unternehmenssteuerung* 20 (3), S. 285–293. DOI: <https://doi.org/10.1007/s00187-009-0078-3>.
- Verheul, Ingeborg et al. (2019). *Data Stewardship on the map: A study of tasks and roles in Dutch research institutes*. DOI: <https://doi.org/10.5281/zenodo.2669150>.
- Versweyveld, Leslie (2016). *We need 500.000 respected data stewards to operate the European Open Science Cloud*. URL: <http://e-irg.eu/news-blog/-/blogs/we-need-500-000-respected-data-stewards-to-operate-the-european-open-science-cloud> (Abgerufen am 19.12.2018).
- Vogl, Raimund; Rudolph, Dominik; Thoring, Anne (2019). „Bringing Structure to Research Data Management Through a Pervasive, Scalable and Sustainable Research Data Infrastructure“. In: *The Art of Structuring*. Hrsg. von Katrin Bergener; Michael Räckers; Armin Stein. Springer International Publishing, 501–512. DOI: https://doi.org/10.1007/978-3-030-06234-7_47.
- Wenger, Etienne (2000). „Communities of Practice and Social Learning Systems“. In: *Organization* 7 (2), S. 225–246. DOI: <https://doi.org/10.1177/135050840072002>.
- Whyte, Angus et al. (2017). *Skills landscape analysis and competence model*. Deliverable 7.1. EOSCpilot. URL: <https://eoscipilot.eu/sites/default/files/eoscipilot-d7.1.pdf> (Abgerufen am 01.05.2019).
- Whyte, Angus et al. (2018). *Skills and capability framework*. Deliverable 7.3. EOSCpilot. URL: <https://eoscipilot.eu/sites/default/files/eoscipilot-d7.3.pdf> (Abgerufen am 11.05.2019).
- Whyte, Angus et al. (2019). *Strategy for Sustainable Development of Skills and Capabilities*. Deliverable 7.5. EOSCpilot. URL: <https://eoscipilot.eu/sites/default/files/eoscipilot-d7.5-v1.1.pdf> (Abgerufen am 01.05.2019).
- Wikström, Ewa (2008). „Boundary work as inner and outer dialogue: dieticians in Sweden“. In: *Qualitative Research in Organizations and Management: An International Journal* 3 (1), S. 59–77. DOI: <https://doi.org/10.1108/17465640810870391>.
- Wilkinson, Mark D. et al. (2016). „The FAIR Guiding Principles for scientific data management and stewardship“. In: *Scientific Data* 3, S. 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.
- Williams, Paul (2002). „The Competent Boundary Spanner“. In: *Public Administration* 80 (1), S. 103–124. DOI: <https://doi.org/10.1111/1467-9299.00296>.
- (2013). „We are all boundary spanners now?“ In: *International Journal of Public Sector Management* 26 (1), S. 17–32. DOI: <https://doi.org/10.1108/09513551311293417>.
- Worrell, Richard; Appleby, Michael C. (2000). „Stewardship of Natural Resources: Definition, Ethical and Practical Aspects“. In: *Journal of Agricultural and Environmental Ethics* 12 (3), S. 263–277. DOI: <https://doi.org/10.1023/A:1009534214698>.
- Yu, Holly H. (2017). „The role of academic libraries in research data service (RDS) provision: Opportunities and challenges“. In: *The Electronic Library* 35 (4), S. 783–797. DOI: <https://doi.org/10.1108/EL-10-2016-0233>.
- Zdunczyk, Kasia (2006). „Human boundary objects—fact or fiction“. In: *OLKC 2006 Conference Proceedings* (University of Warwick, Coventry, 20. März 2006–22. März 2006),

S. 20–22. URL: <https://pdfs.semanticscholar.org/6866/77b7d9407cf9d10e8950ff90ebc781ea02ab.pdf> (Abgerufen am 03.05.2019).

Zuiderwijk, Anneke; Janssen, Marijn (2014). „Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View“. In: *Public Administration and Information Technology*. Springer New York, S. 115–135. DOI: https://doi.org/10.1007/978-1-4614-9563-5_8.

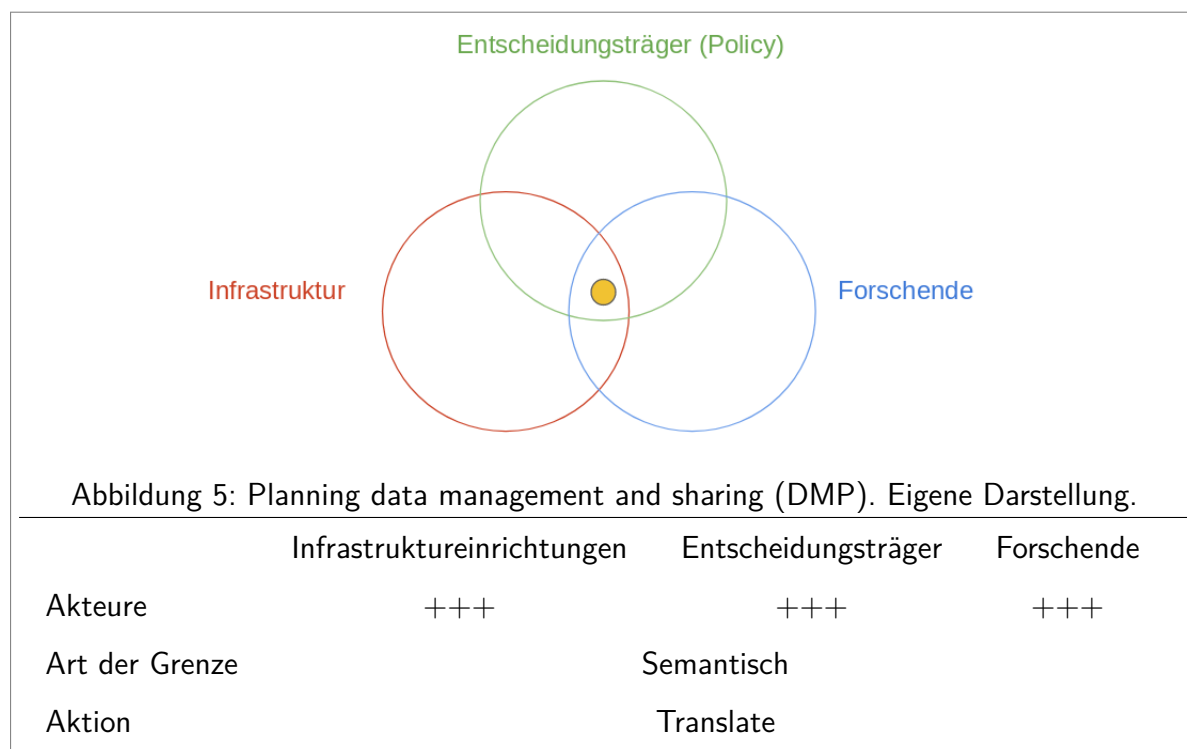
Anhang

Anhang A Tätigkeiten und Akteure

In diesem Anhang wird die Verortung der im EOSCpilot entwickelten Data Stewardship Tätigkeiten analog zu ihrer Beschreibung in Kapitel 6 schematisch dargestellt. Die Kreise symbolisieren dabei die Sozialen Welten der Akteure (Infrastruktureinrichtungen = **rot**, Entscheidungsträger = **grün**, Forschende = **blau**). Der gelbe Kreis bezeichnet die Tätigkeit und ihre Verortung innerhalb der Sozialen Welt und relativ zu den anderen Akteuren. Tätigkeiten können an Überschneidungsgrenzen der Welten (Boundaries) angesiedelt sein oder lassen sich nur innerhalb einer Welt lokalisieren. Die Positionierung des gelben Kreises zeigt entsprechend die „Entfernung“ der Tätigkeit von den anderen Akteuren. Beispielsweise können Tätigkeiten zwar in der Nähe anderer Akteure liegen, jedoch die Boundary nicht ganz erreichen.

Unterhalb der schematischen Darstellungen gibt die Tabelle mit Hilfe von + Zeichen die Intensität der Beteiligung der jeweiligen Akteure an. + bedeutet einfache Beteiligung, ++ starke Beteiligung und +++ sehr intensive Beteiligung. In Klammern gesetzte Zeichen zeigen Überlegungen zu eventuellen Beteiligungen an. Die Intensität der Beteiligung ist nicht gleichzusetzen mit ihrer Verortung innerhalb der Sozialen Welt. Beispielsweise kann eine Tätigkeit nah zu einer Grenze einer anderen Akteursgruppe liegen und trotzdem eine sehr intensive Beteiligung der Akteursgruppe, in deren Welt sie eingeordnet ist, mit sich bringen. Auch Tätigkeiten, die sehr nah an den Grenzen liegen, können eine einfache Beteiligung anderer Akteure anzeigen.

A.1 Plan and design



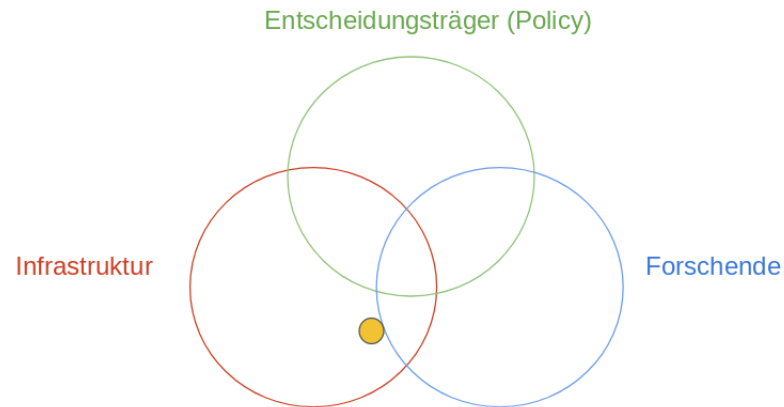


Abbildung 6: Open data model and database design. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		+
Art der Grenze	Syntaktisch/Semantisch		
Aktion	Transfer/Translate		

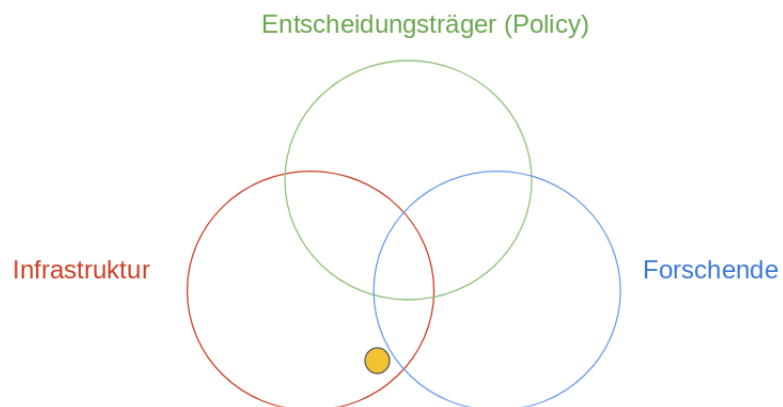


Abbildung 7: Metadata, persistent ID specification. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		+
Art der Grenze	Semantisch		
Aktion	Translate		

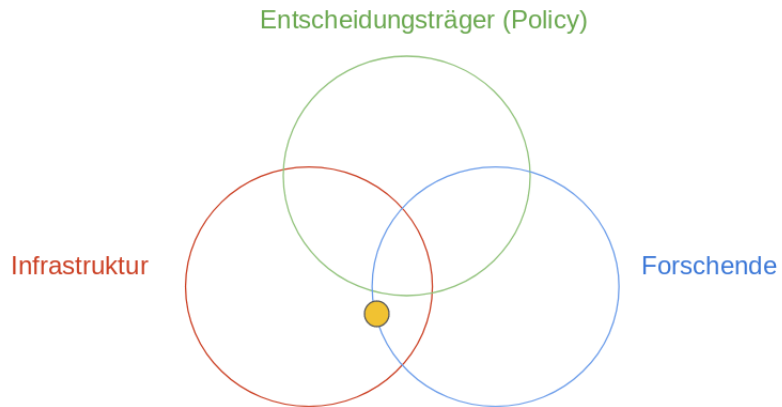


Abbildung 8: Open source software / service requirements. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		++
Art der Grenze	Semantisch/Pragmatisch		
Aktion	Translate/Transform		

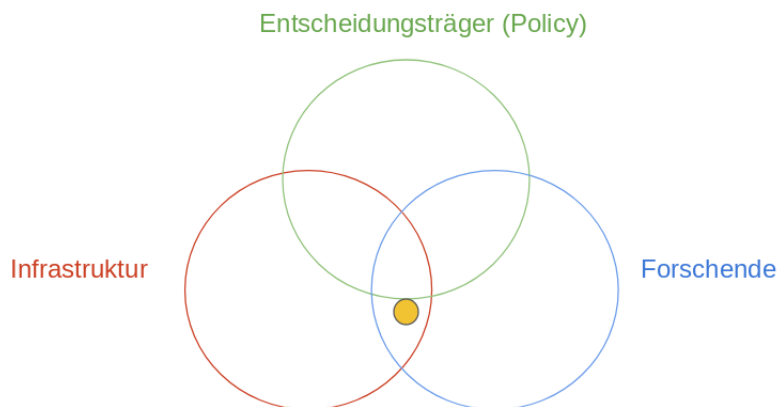
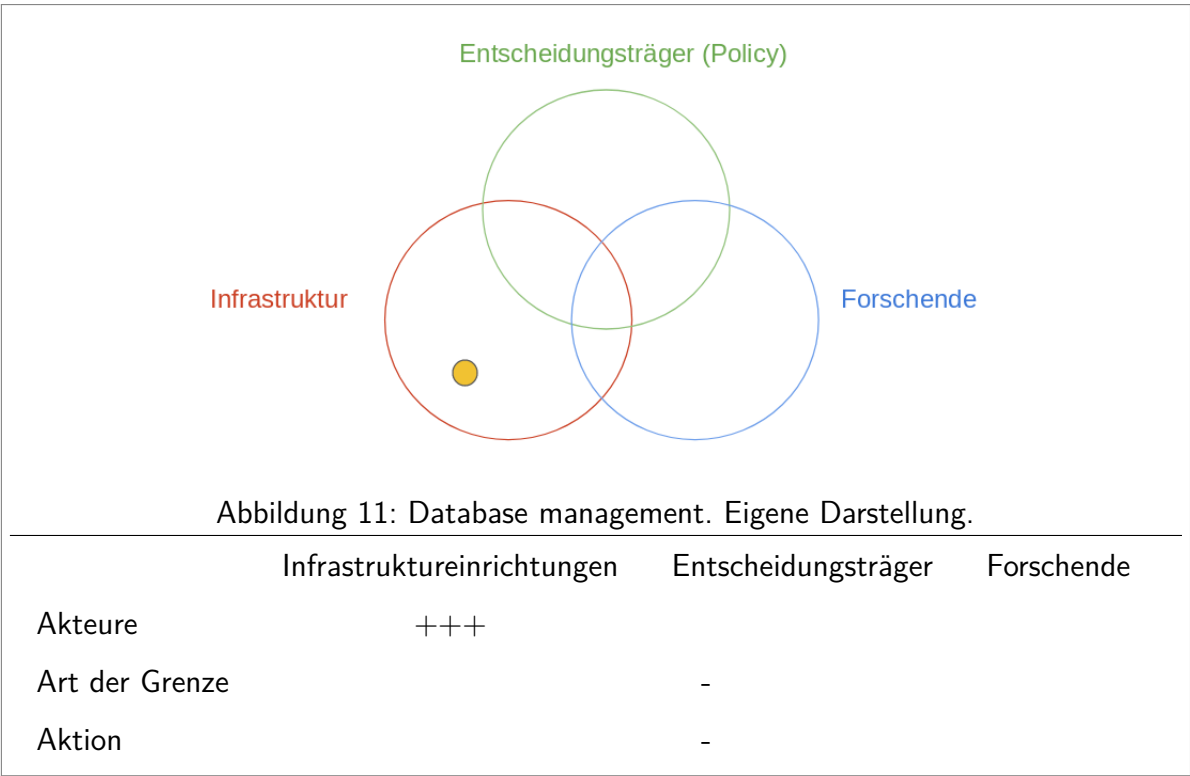
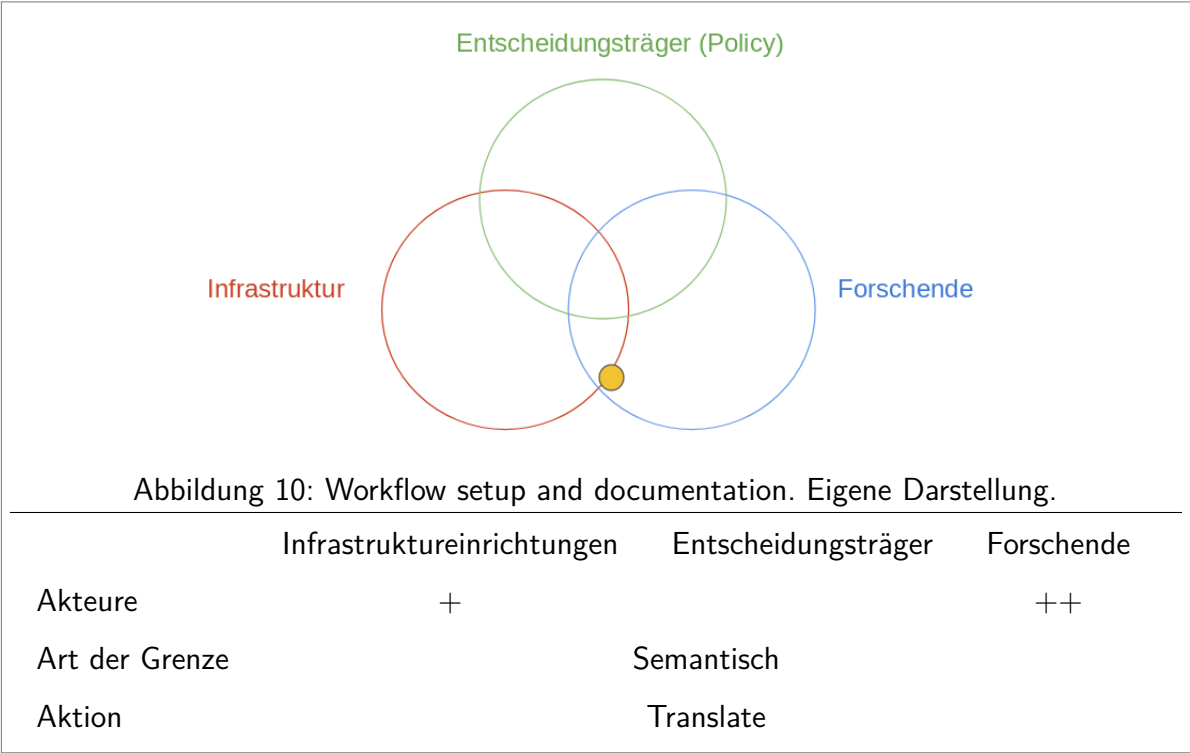


Abbildung 9: Repository and data management platform appraisal. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	++	+	++
Art der Grenze	Semantisch		
Aktion	Translate		

A.2 Capture and process



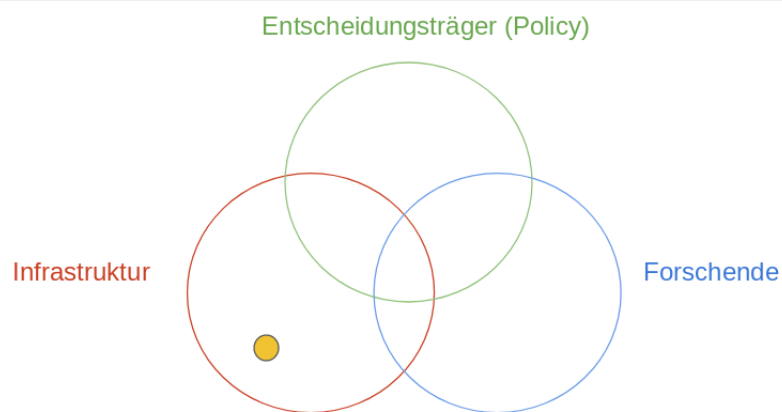


Abbildung 12: Software prototyping. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	

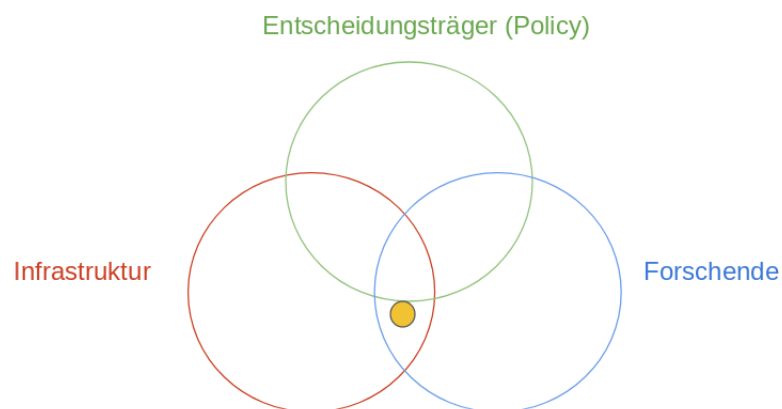
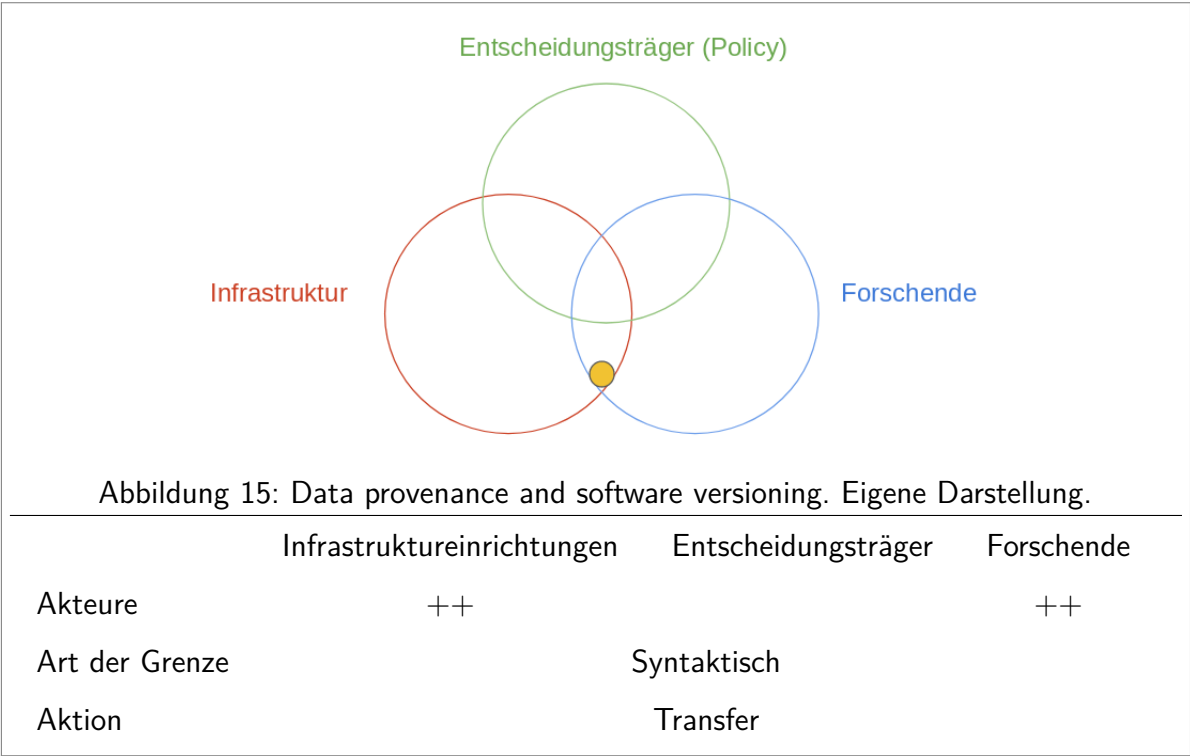
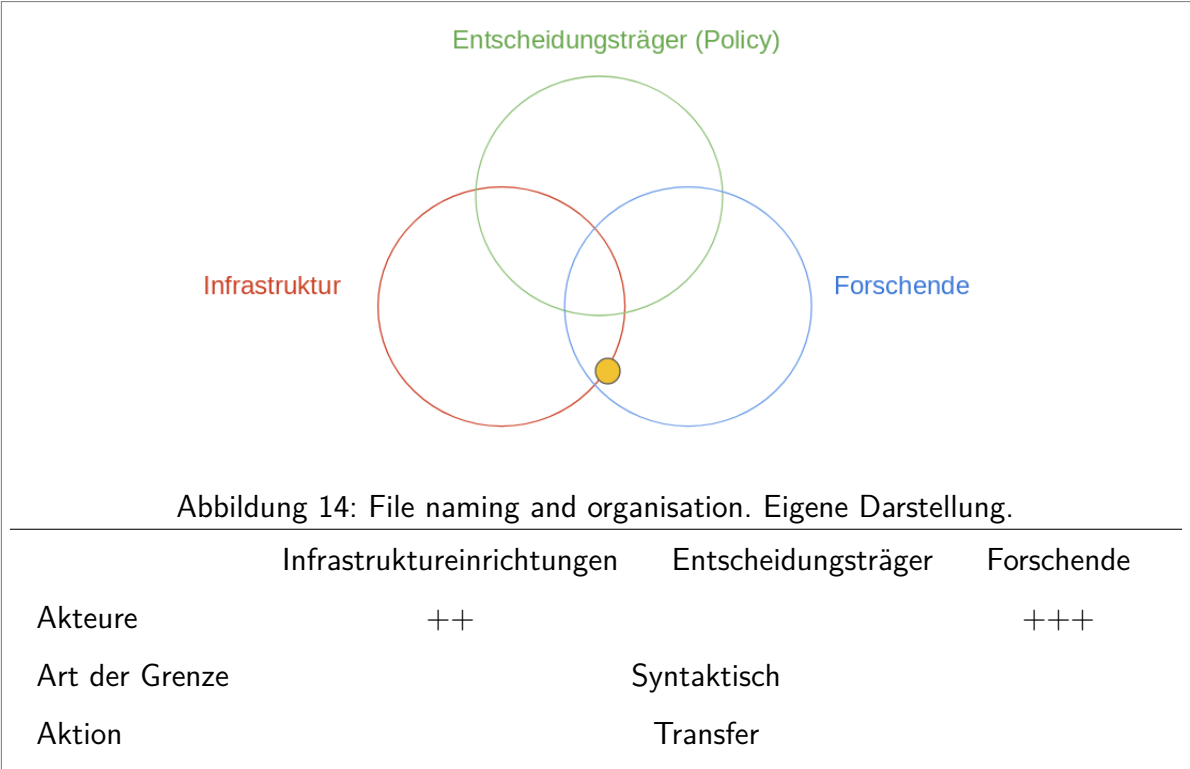
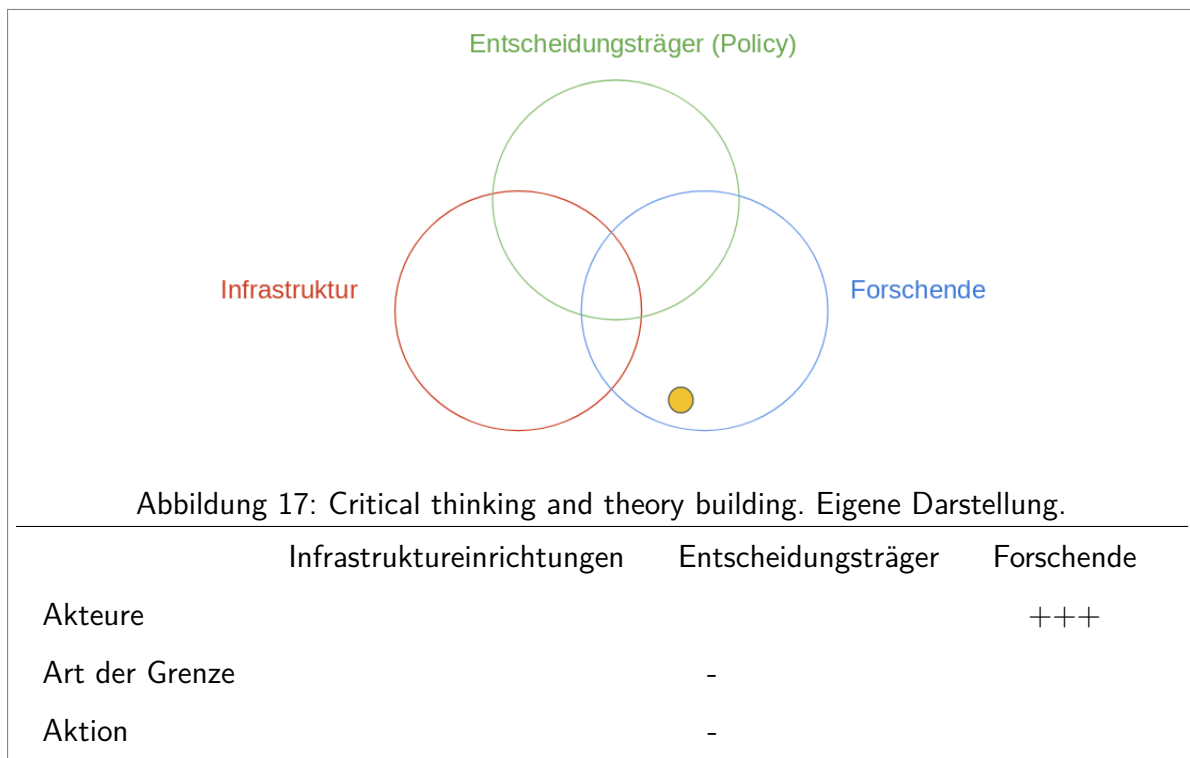
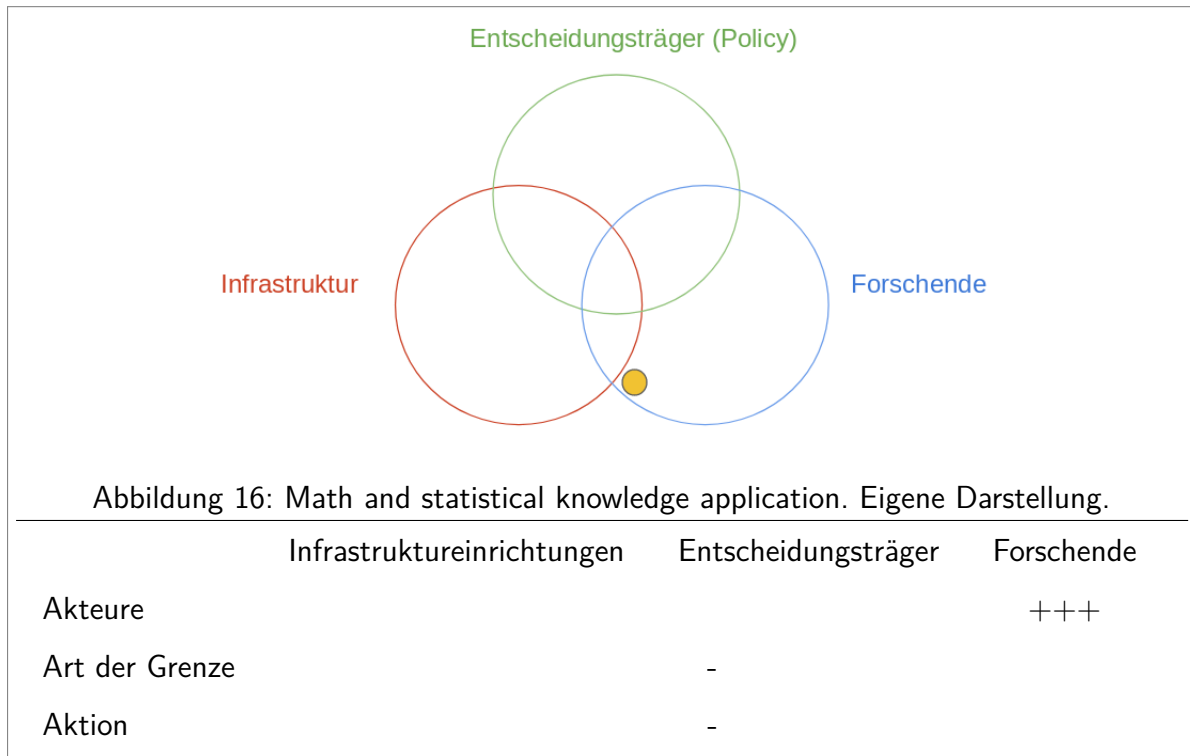


Abbildung 13: Data collection and reuse of open data. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	++	+	++
Art der Grenze	Semantisch/Pragmatisch		
Aktion	Translate/Transform		



A.3 Integrate and analyse



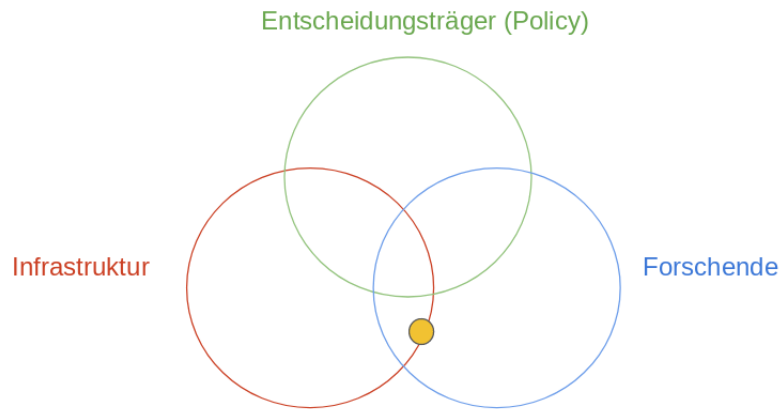


Abbildung 18: Creative problem solving, flexibility. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+		+++
Art der Grenze	Syntaktisch/Semantisch		
Aktion	Transfer/Translate		

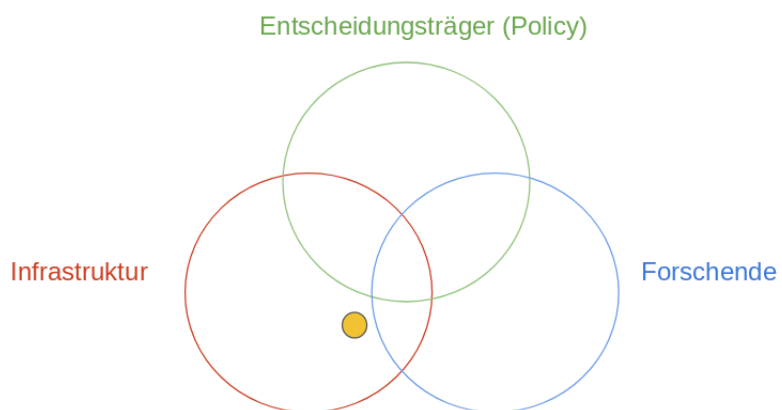
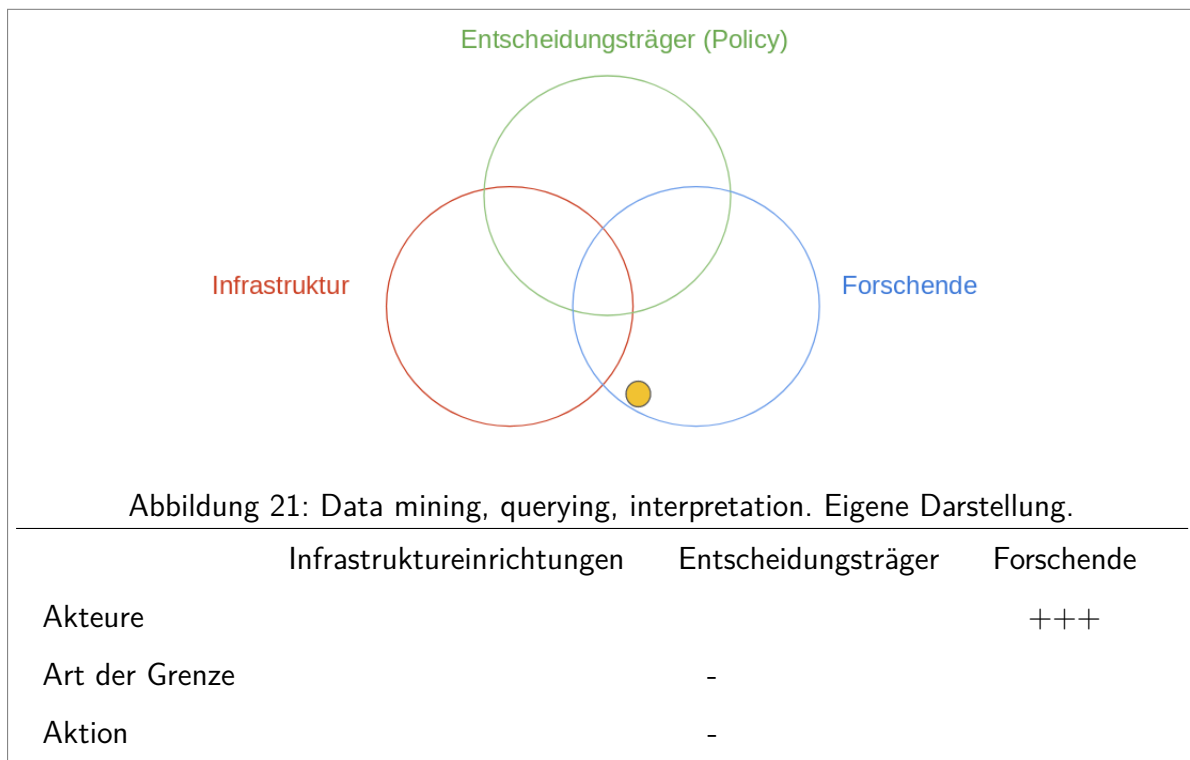
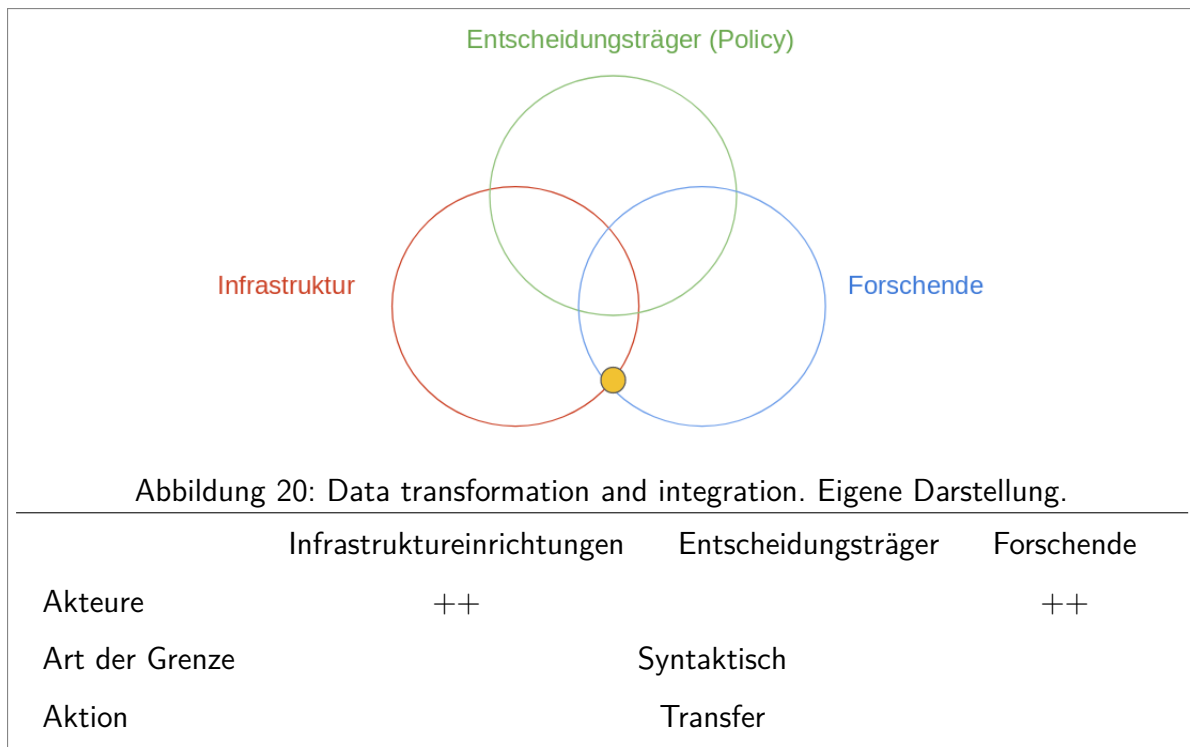


Abbildung 19: Open source software / service development. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		+
Art der Grenze	Semantisch		
Aktion	Translate		



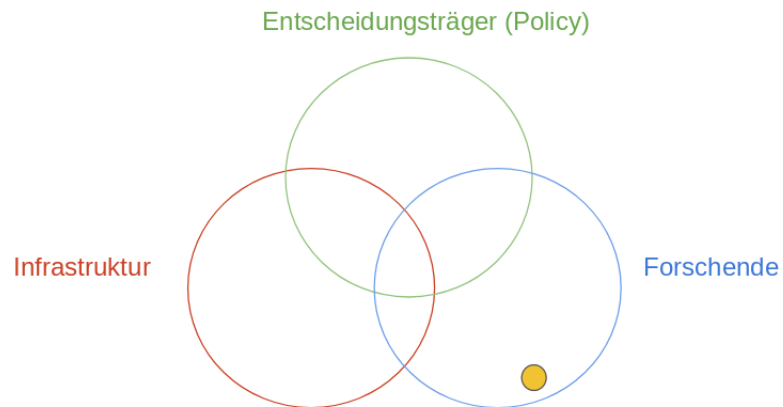


Abbildung 22: Predictive modelling and analytics. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure			+++
Art der Grenze		-	
Aktion		-	

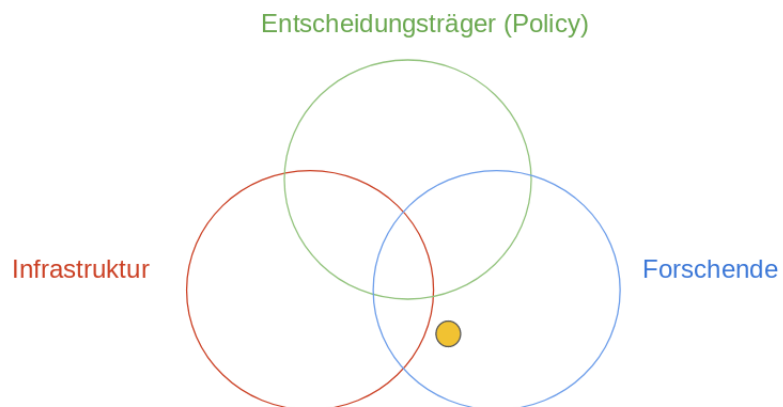
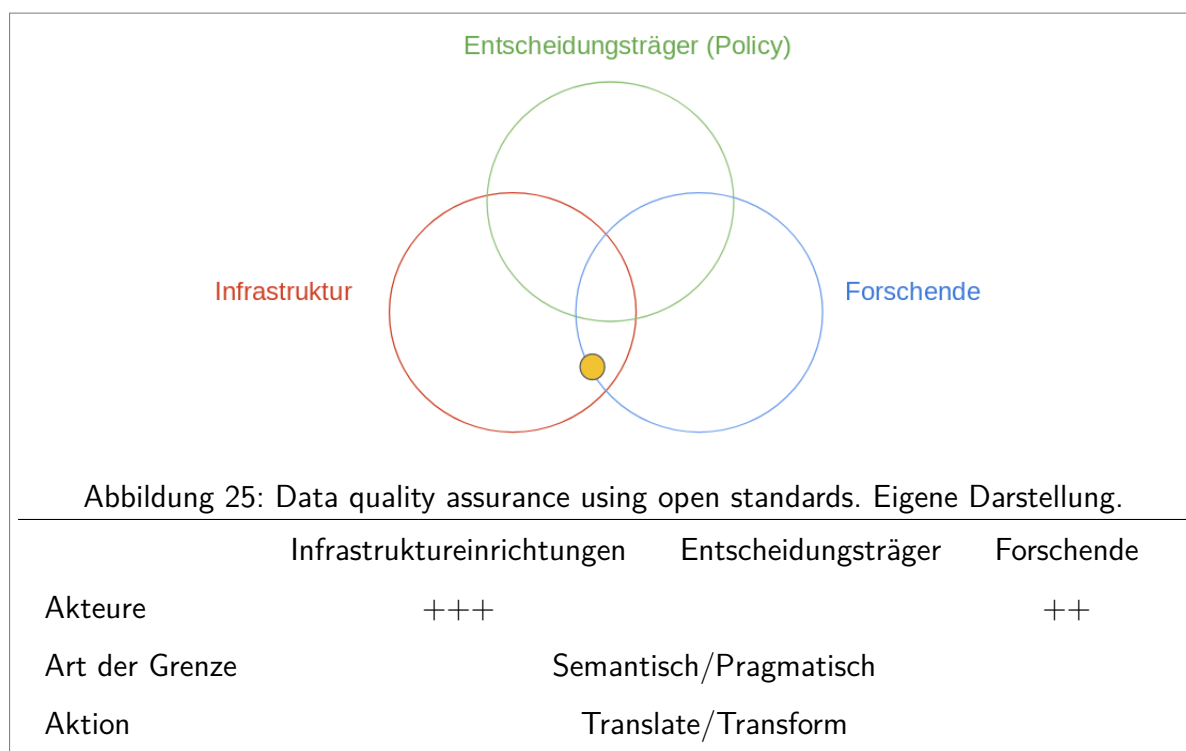
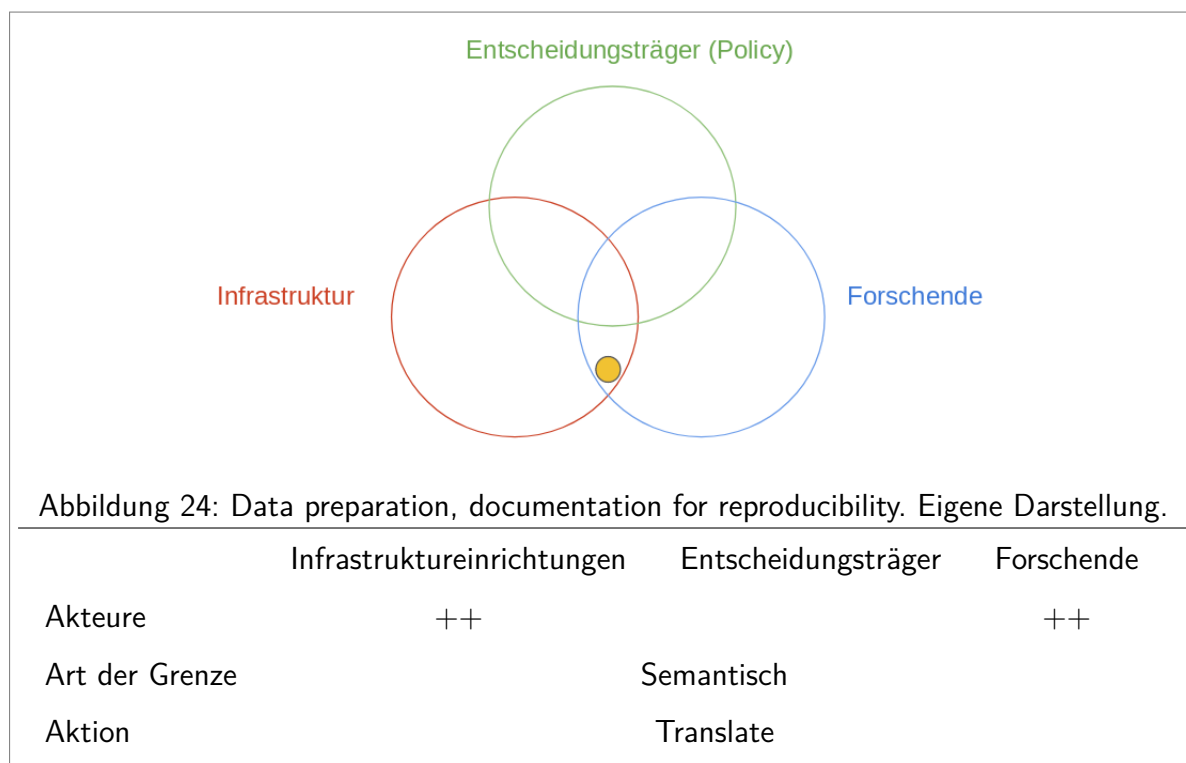


Abbildung 23: Machine learning methods. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure			+++
Art der Grenze		-	
Aktion		-	

A.4 Appraise and preserve



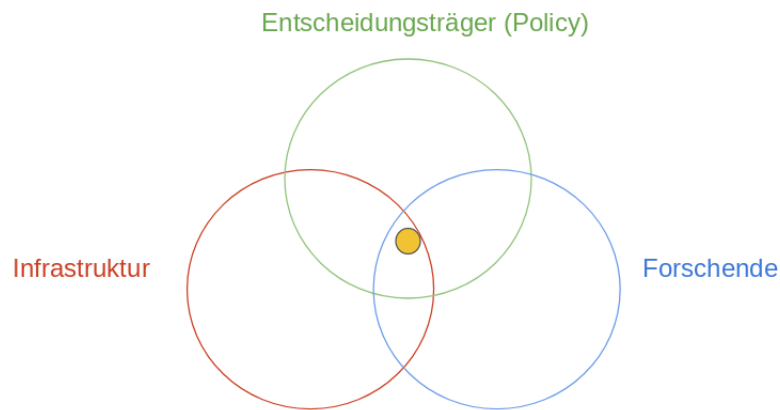


Abbildung 26: Ethical, legal and data policy compliance. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	++	+++	+++
Art der Grenze	Semantisch/Pragmatisch		
Aktion	Translate/Transform		

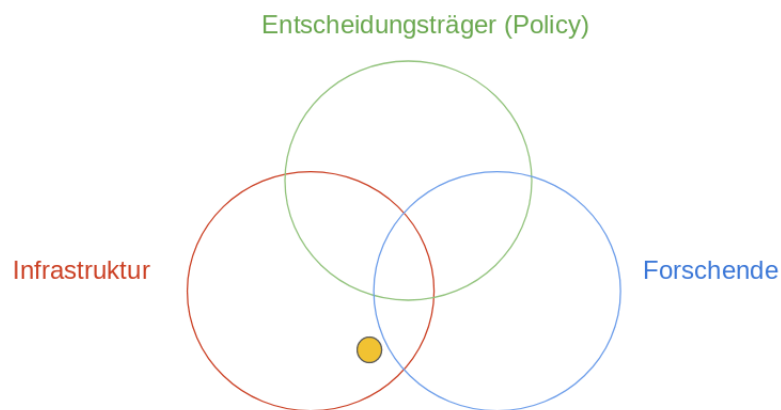


Abbildung 27: Data transfer and long-term storage. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		+
Art der Grenze	Syntaktisch		
Aktion	Transfer		

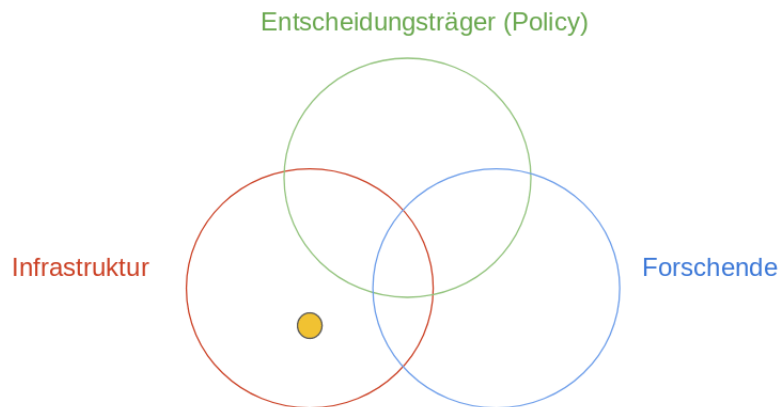


Abbildung 28: Format and media migration. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	

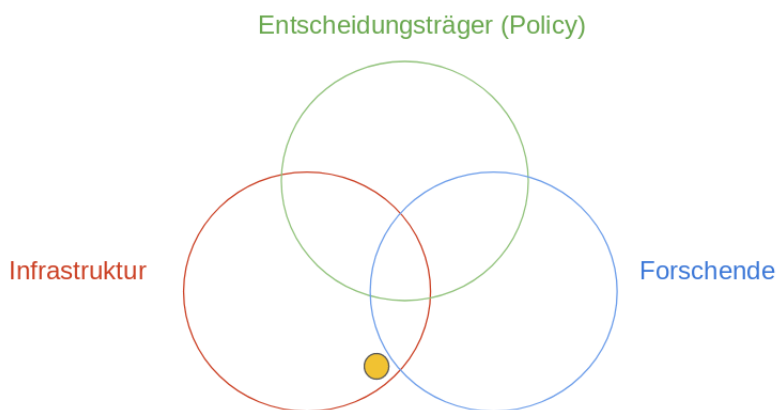
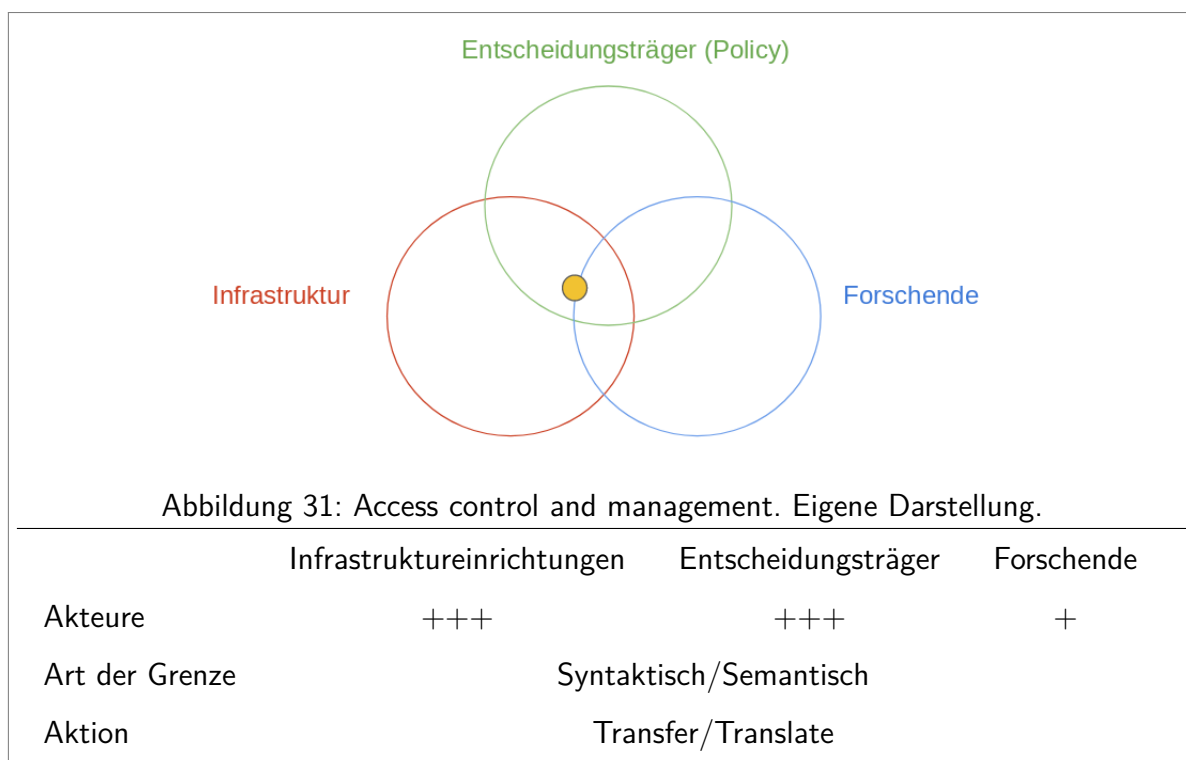
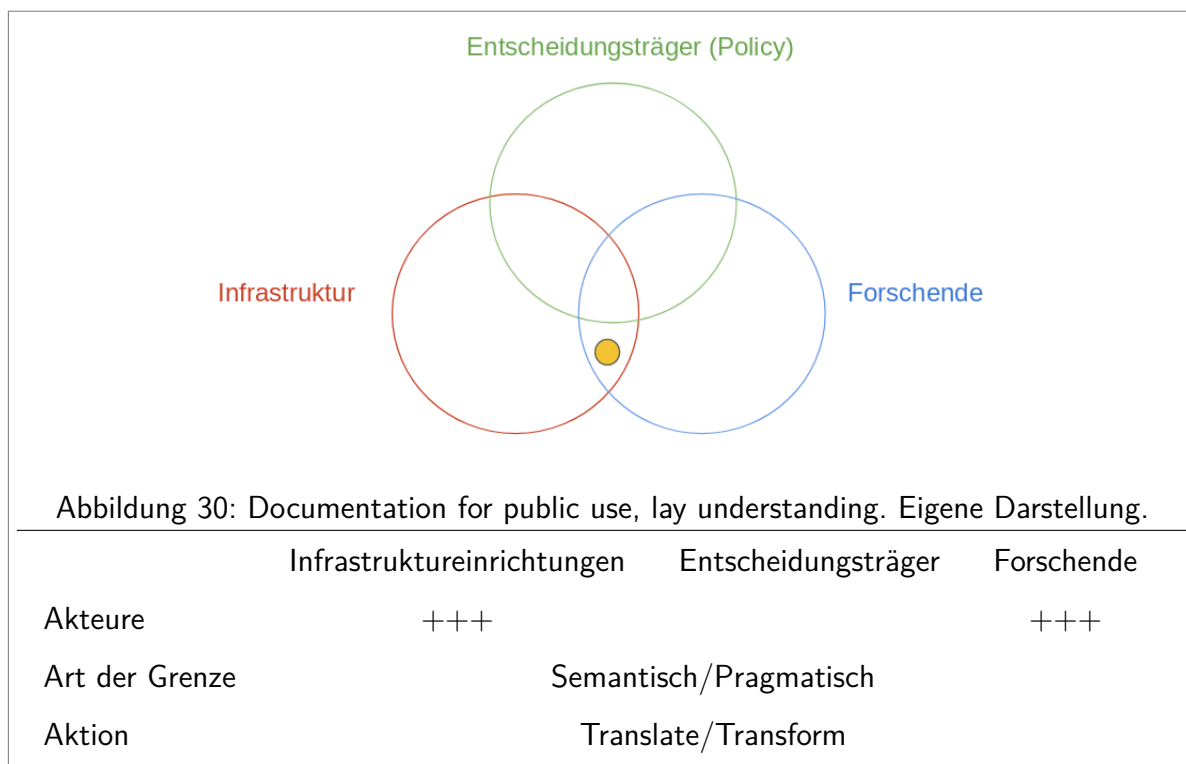
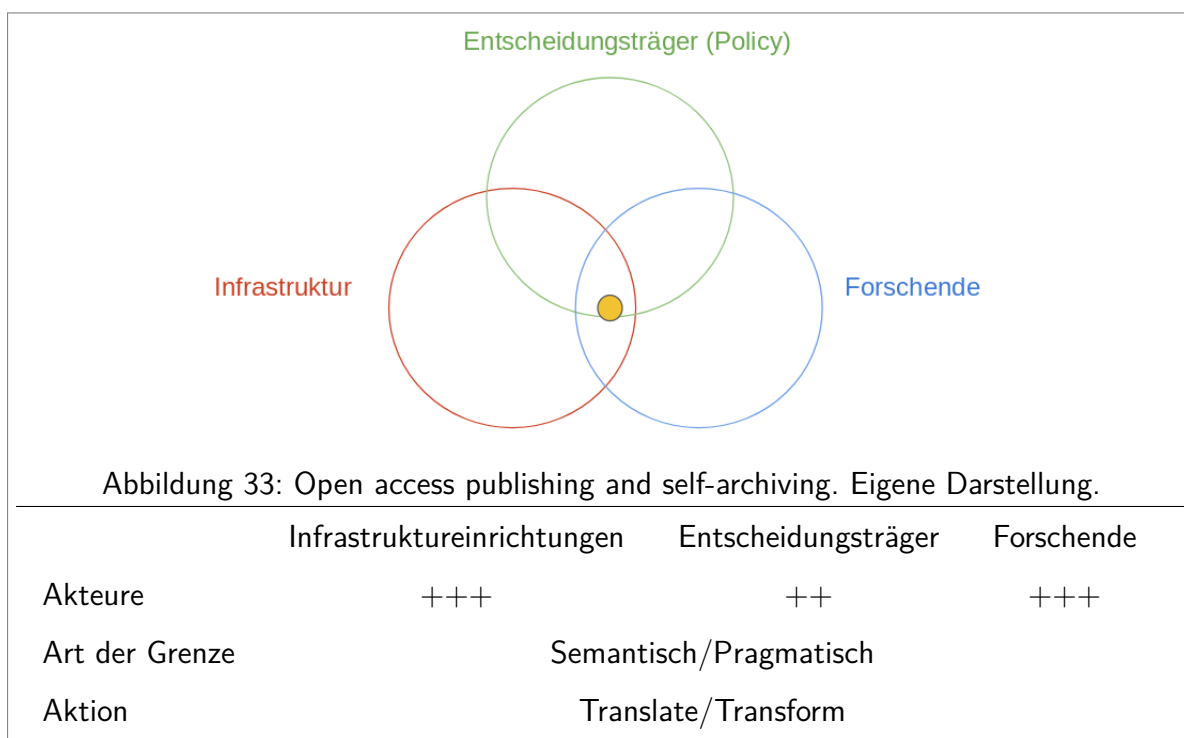
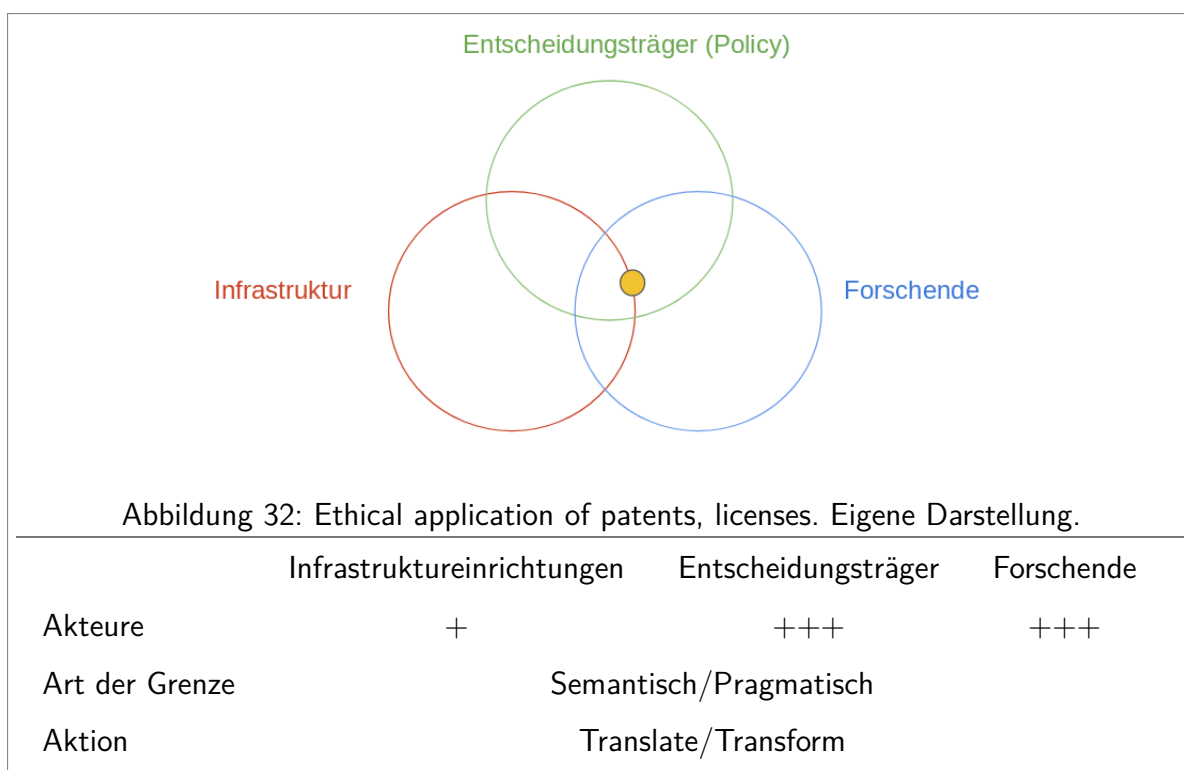


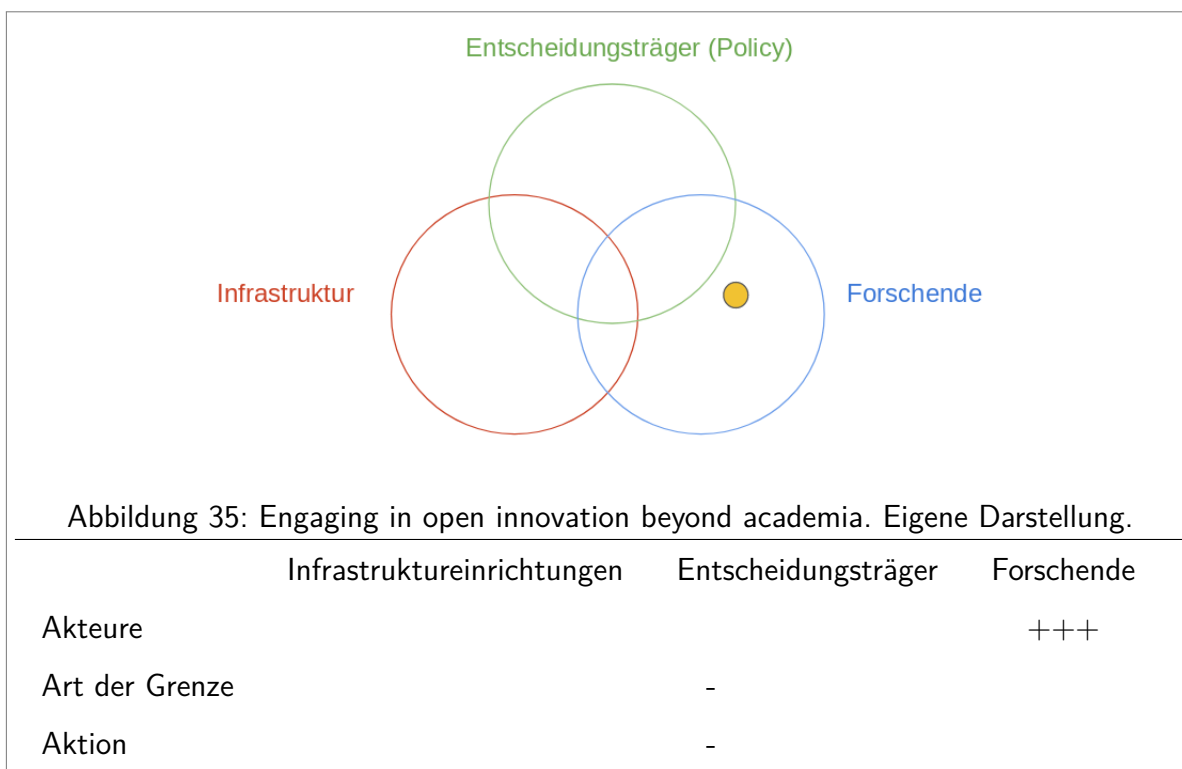
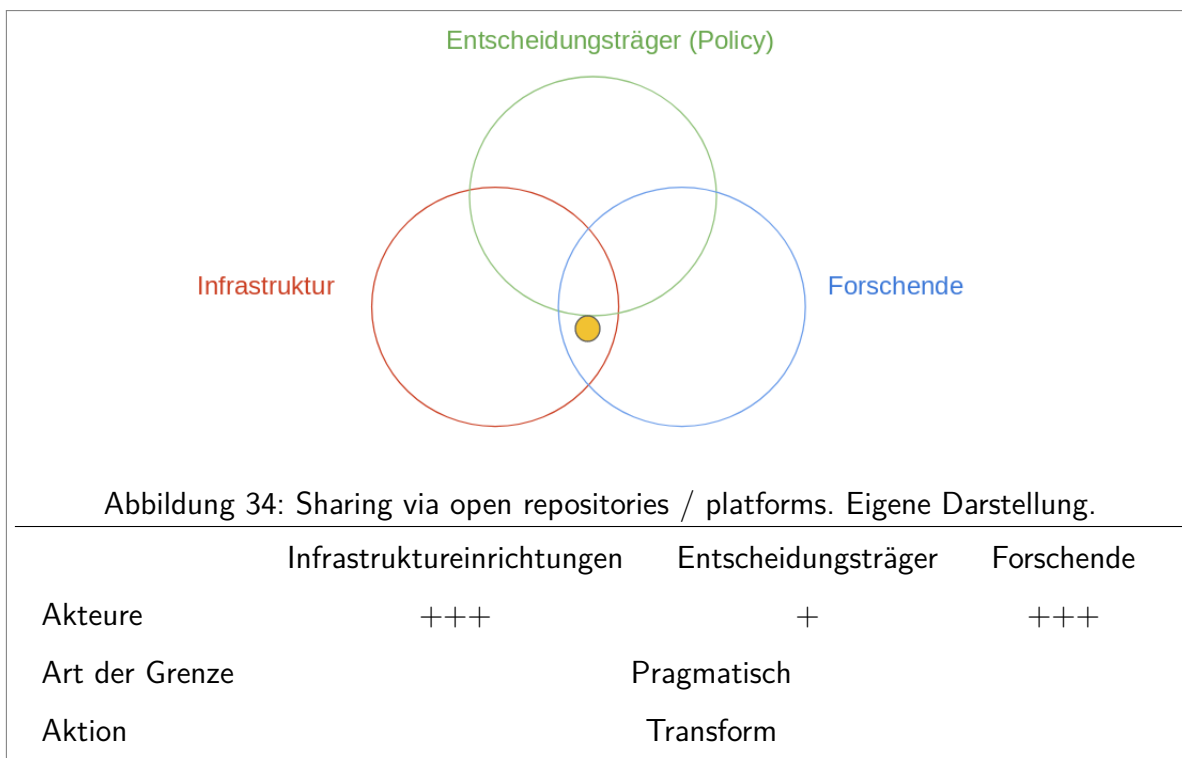
Abbildung 29: Software review and preservation. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		(-)	

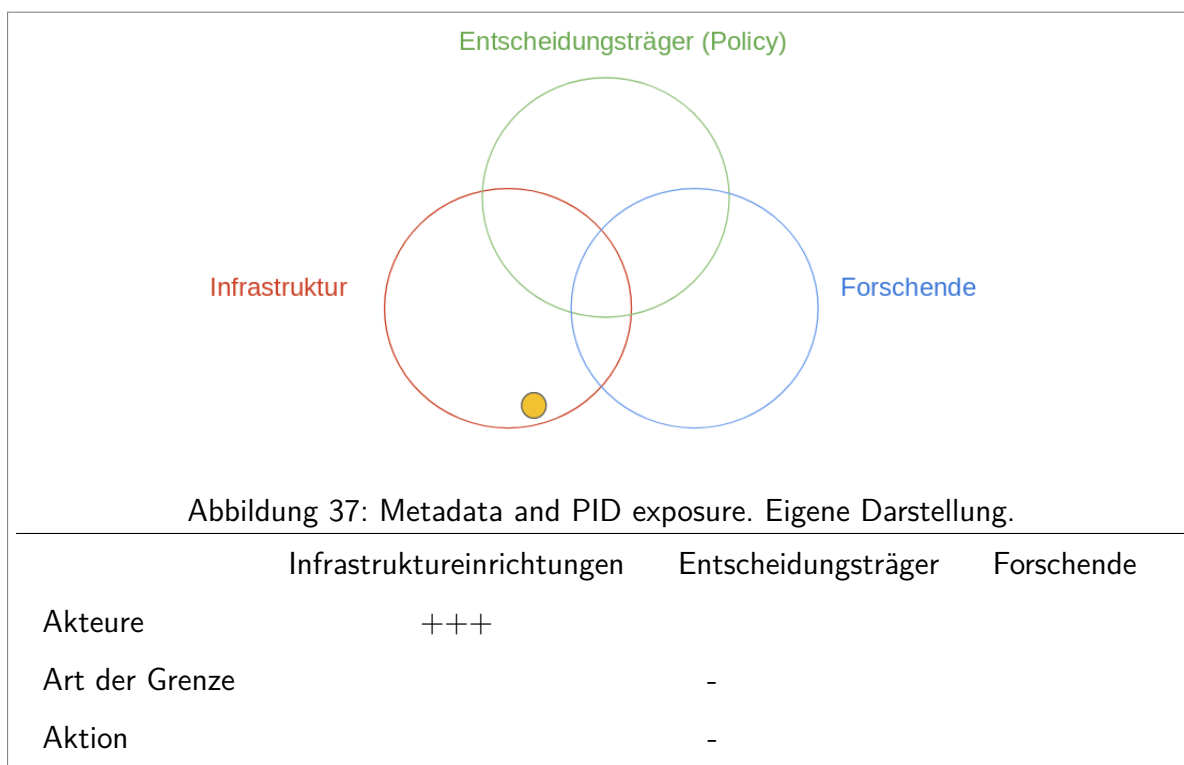
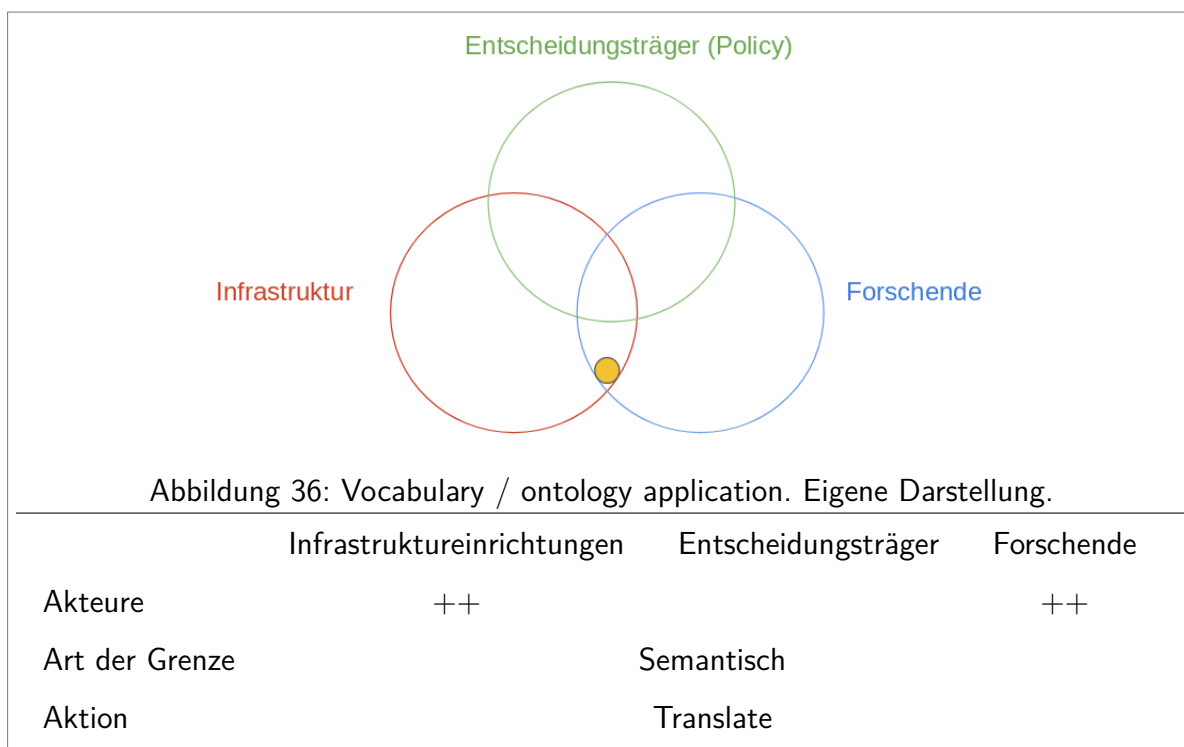
A.5 Publish and release

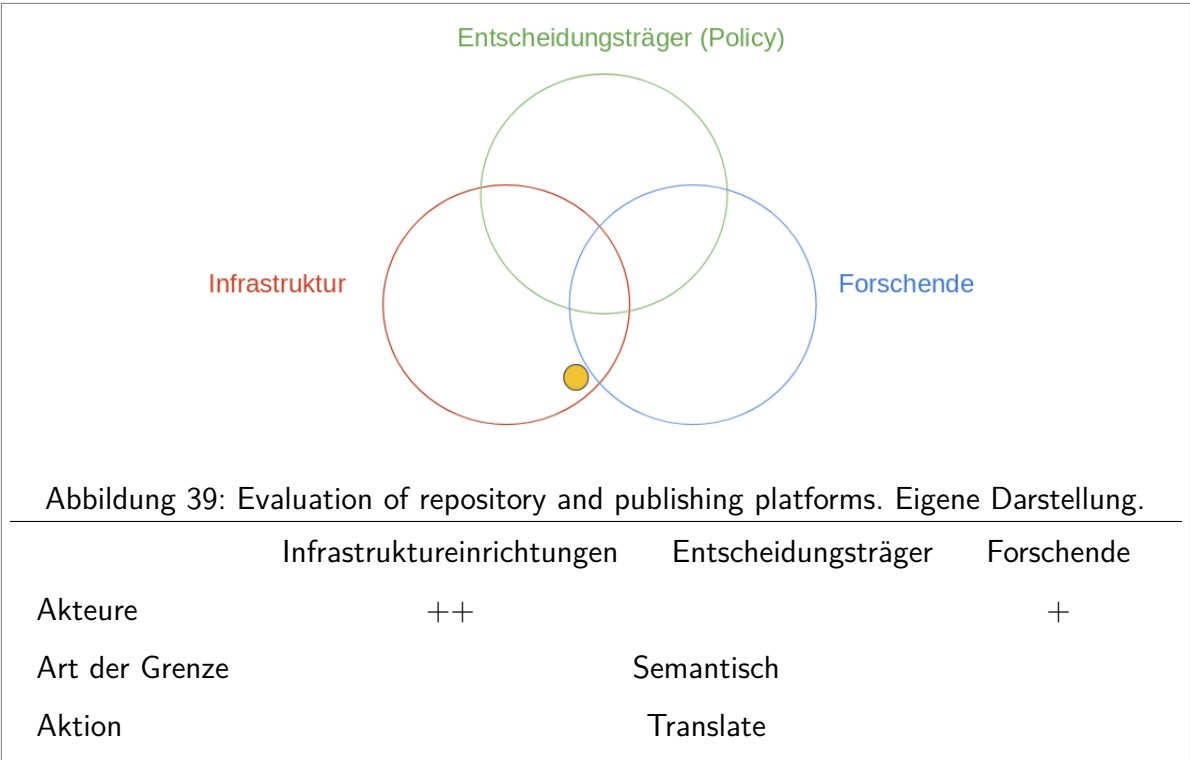
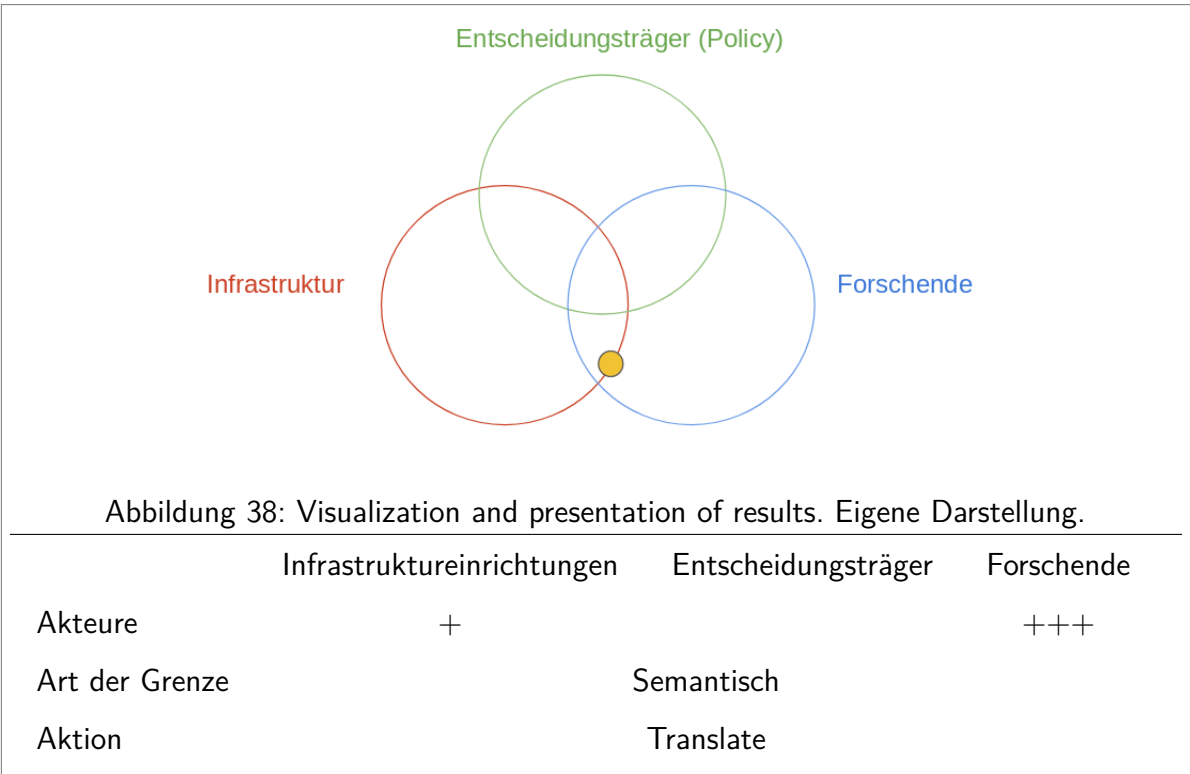


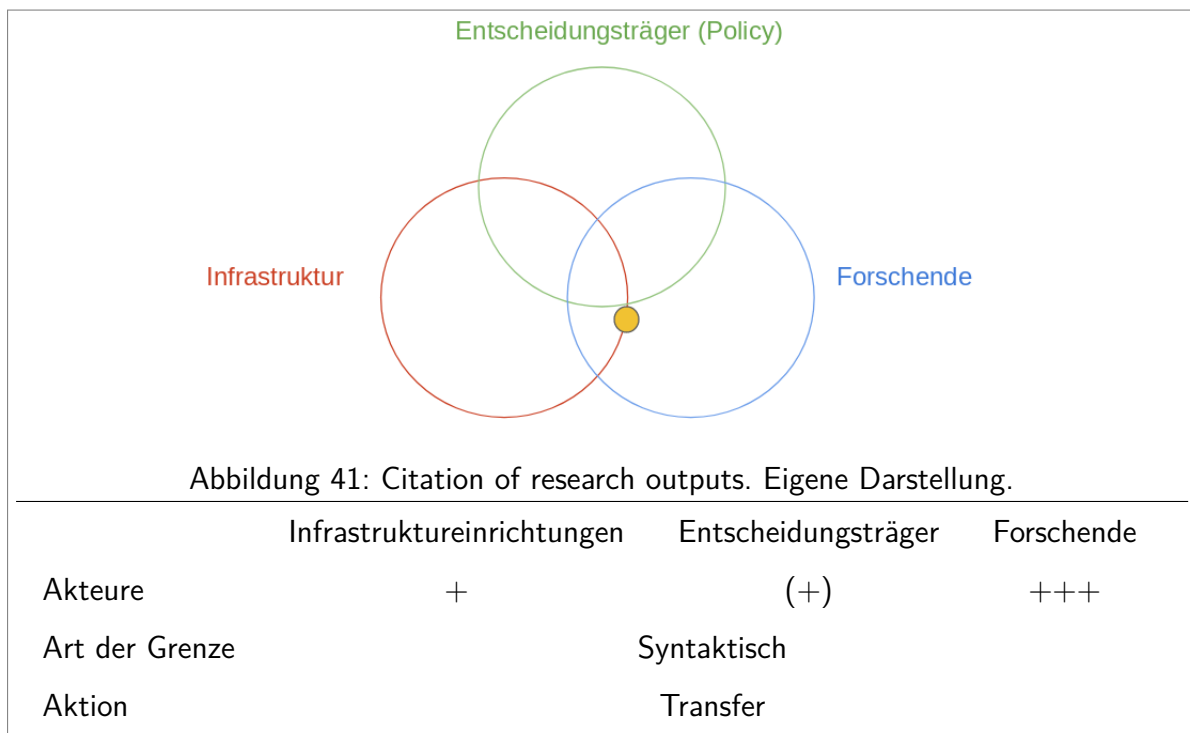
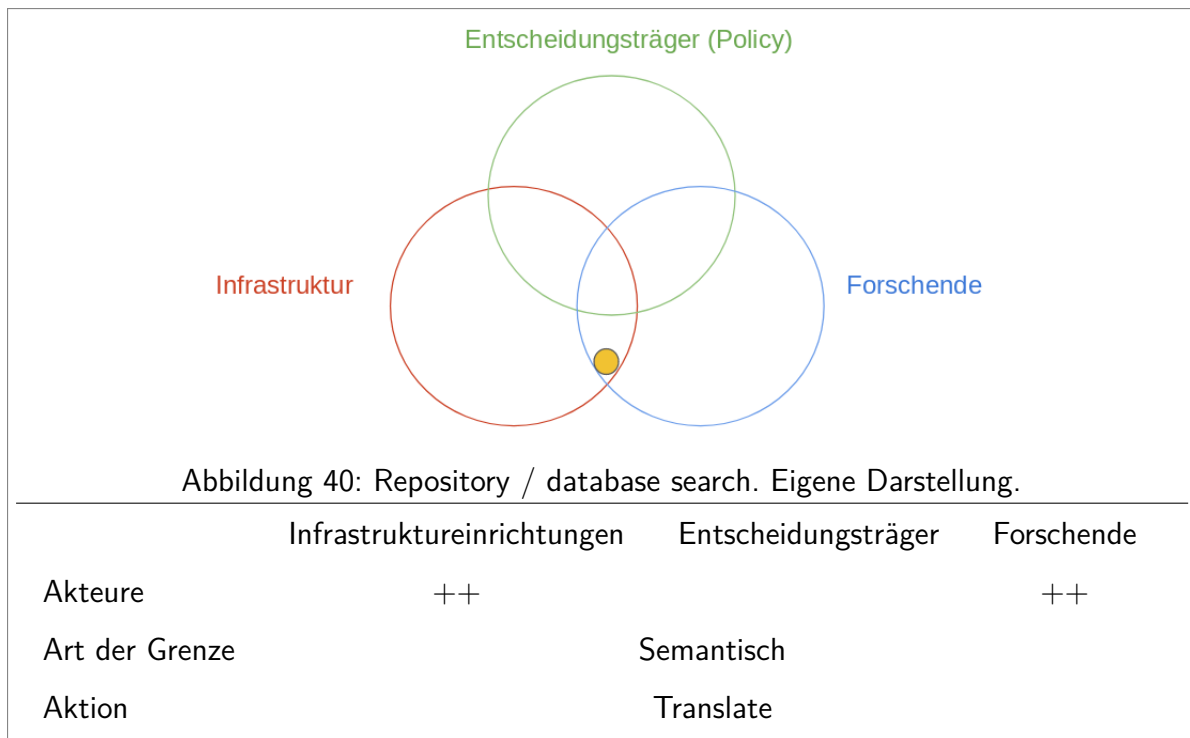




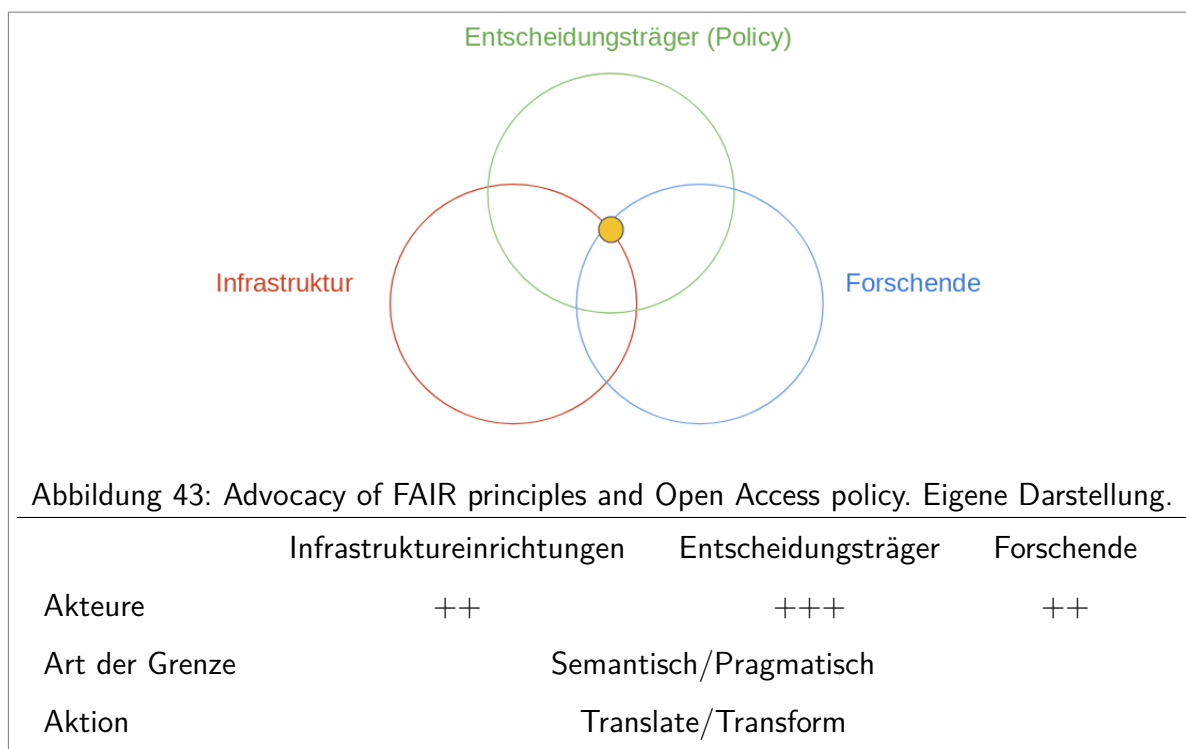
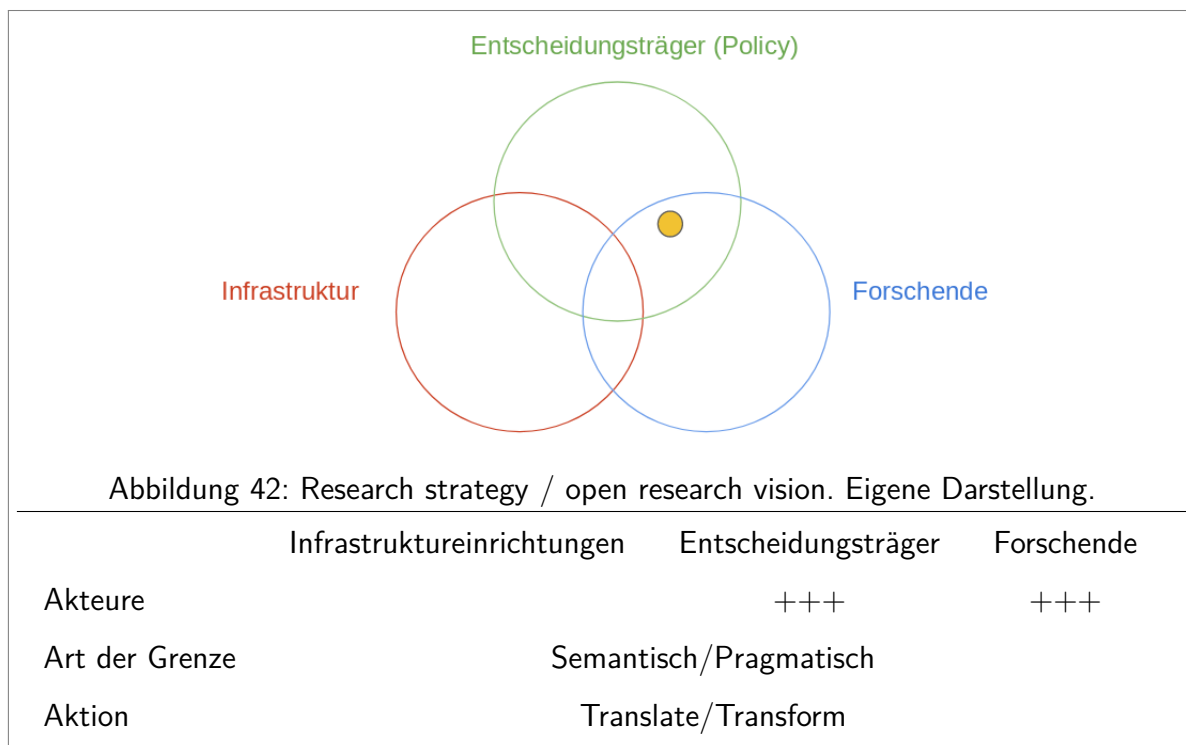
A.6 Expose and discover

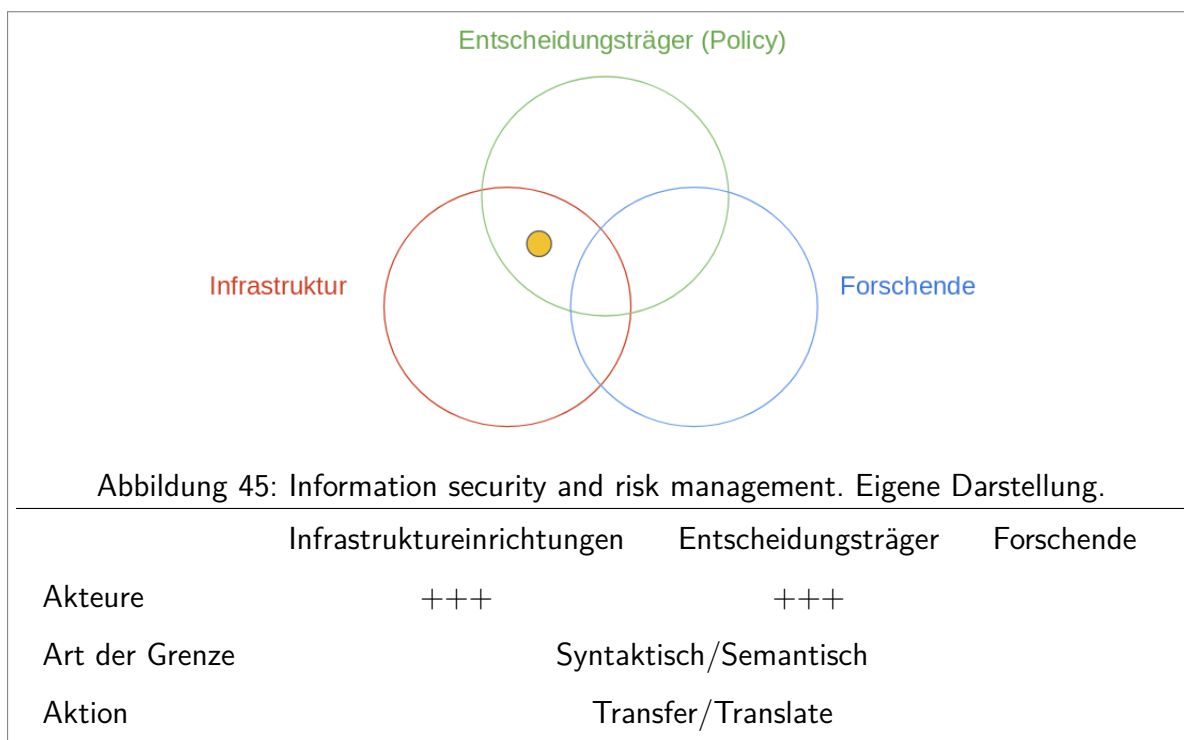
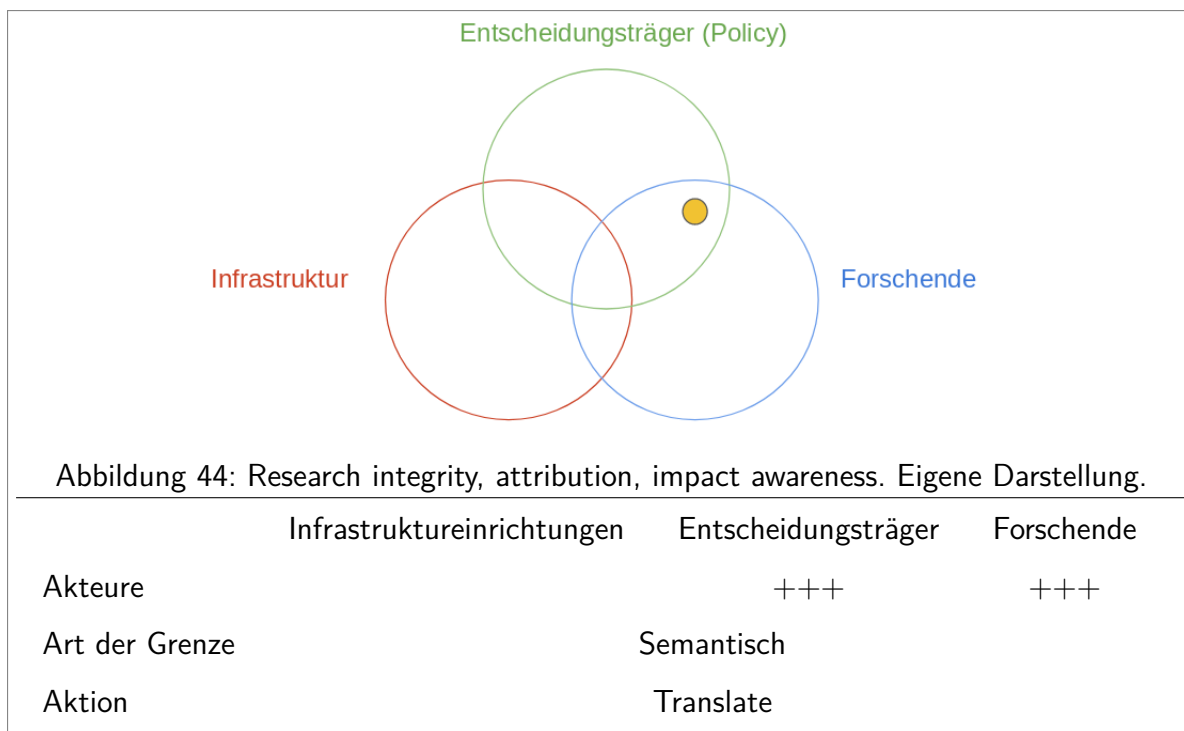


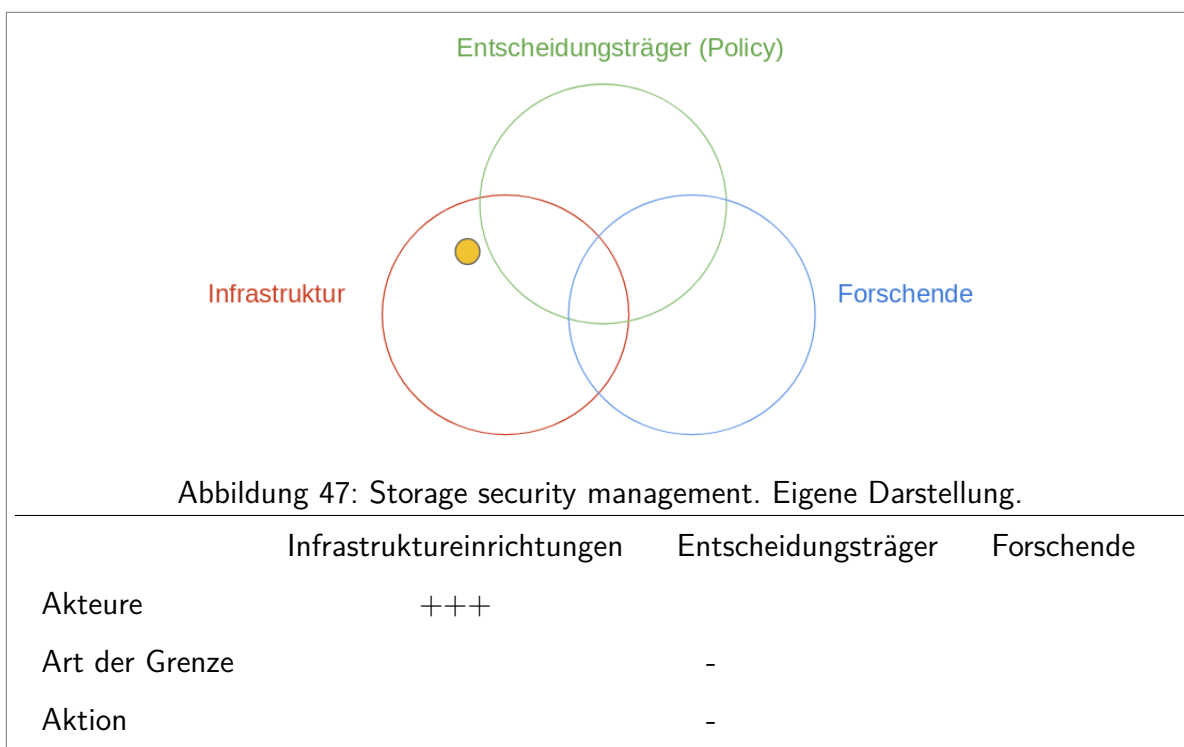
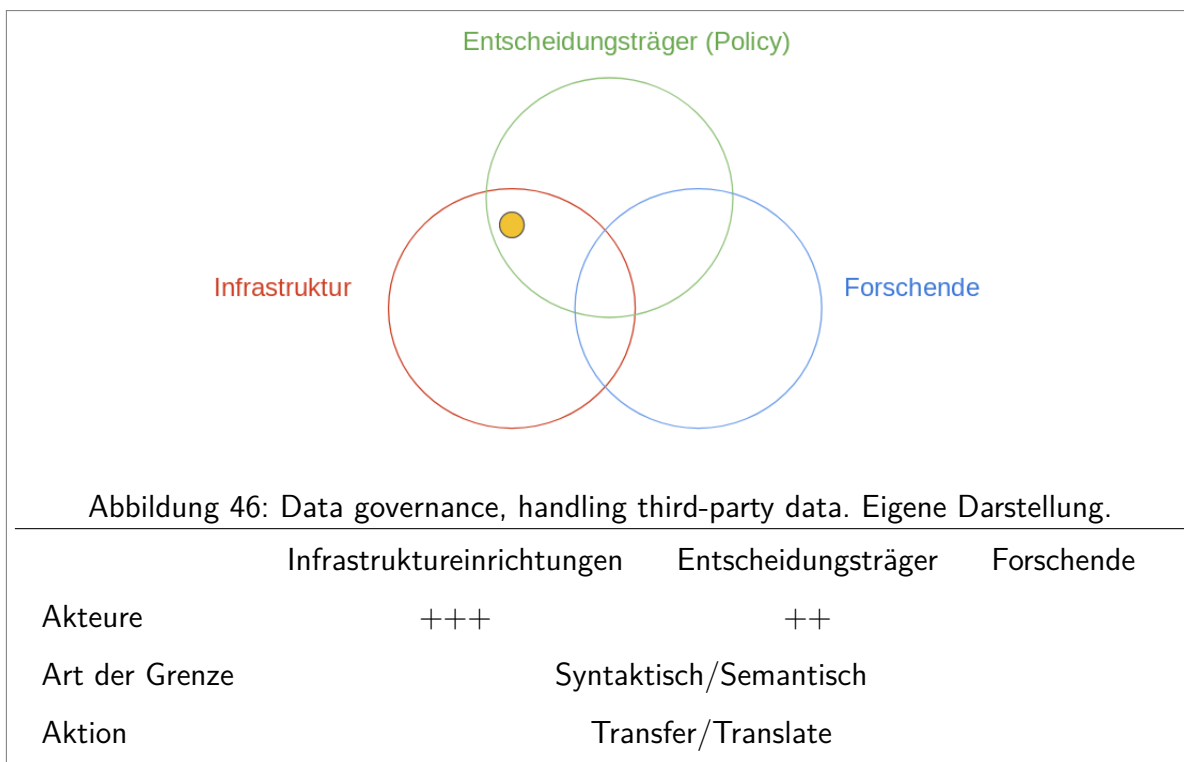


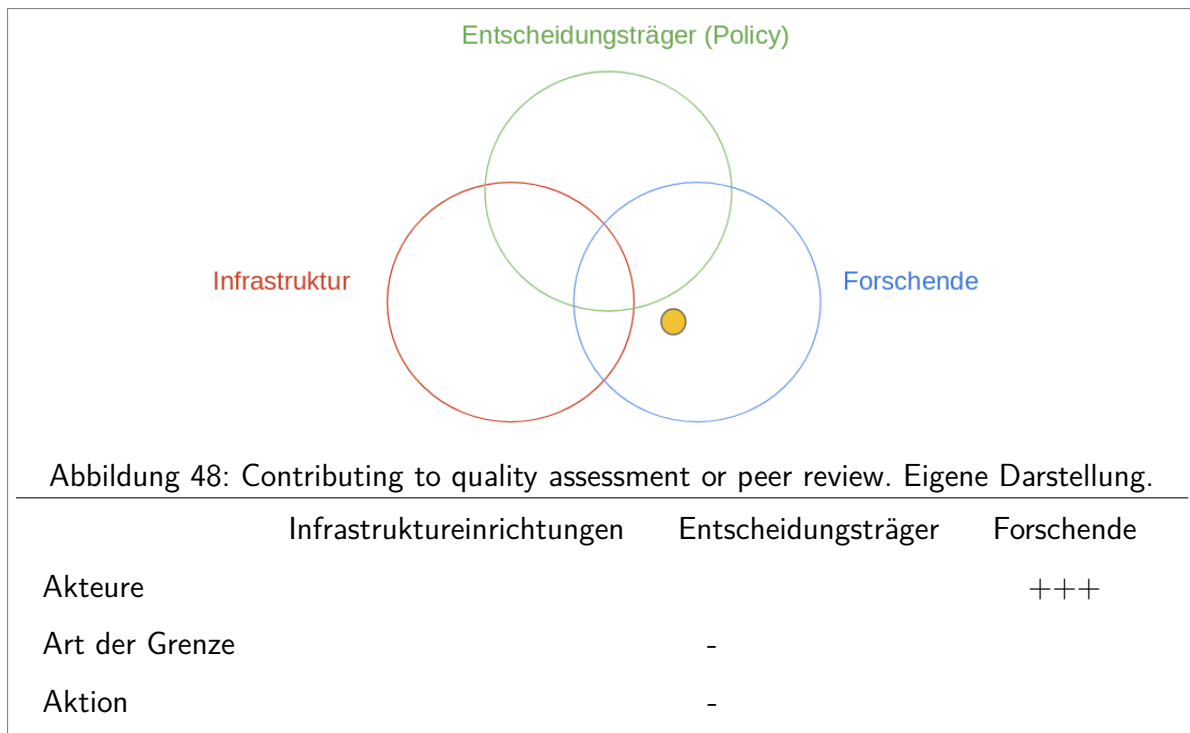


A.7 Govern and assess

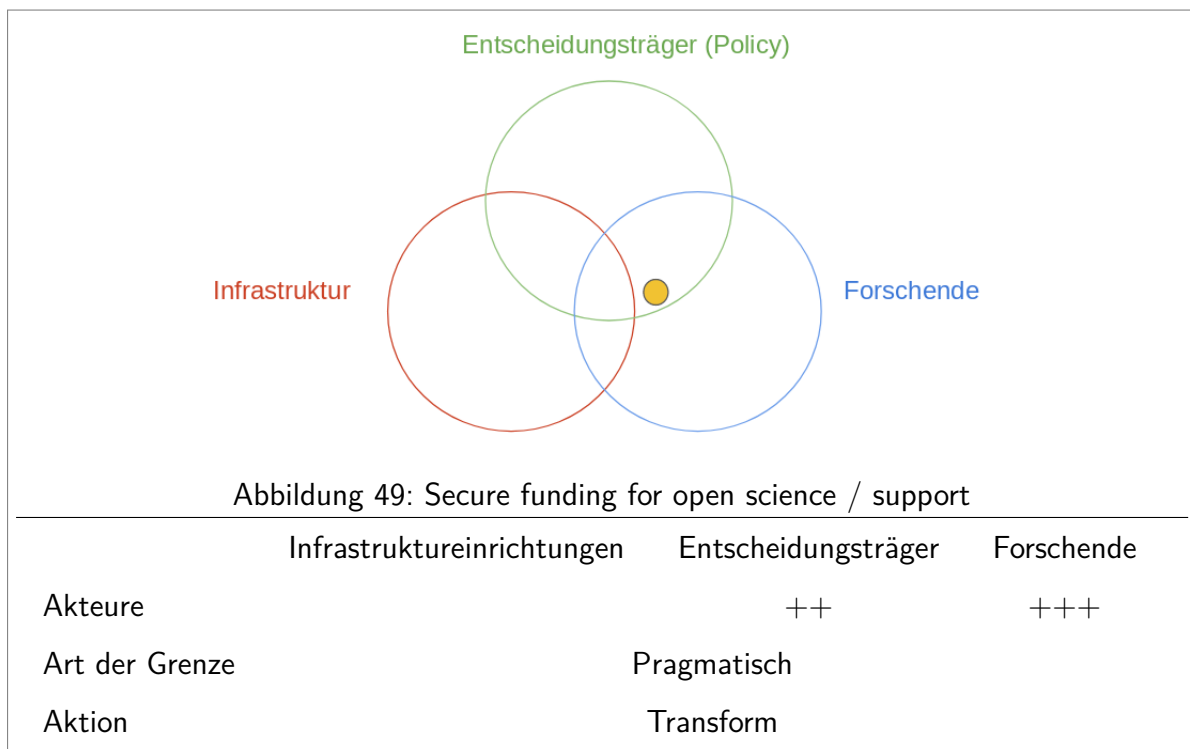








A.8 Scope and resource



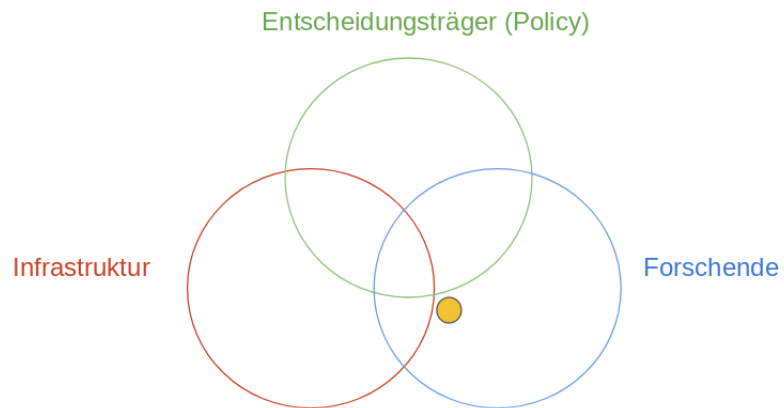


Abbildung 50: Perseverance delivering diverse open research. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	(+)	(+)	+++
Art der Grenze		-	
Aktion		-	

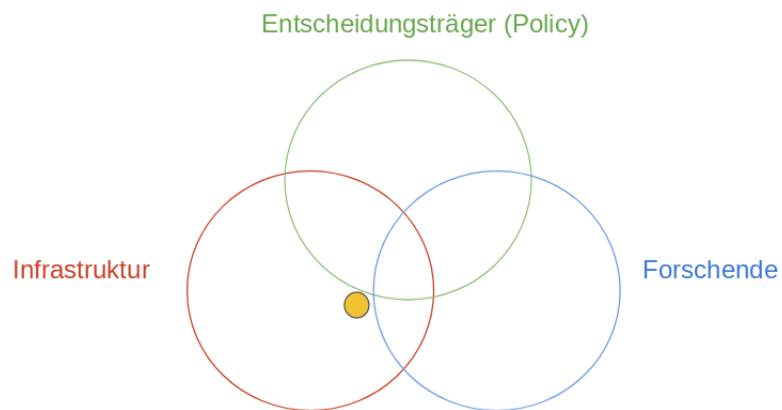


Abbildung 51: Service level management. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++	+	+
Art der Grenze		Semantisch	
Aktion		Translate	

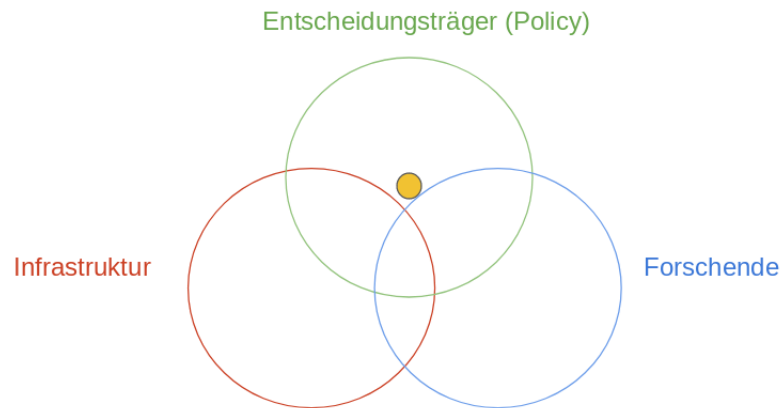


Abbildung 52: Change management. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+	+++	+
Art der Grenze	Pragmatisch		
Aktion	Transform		

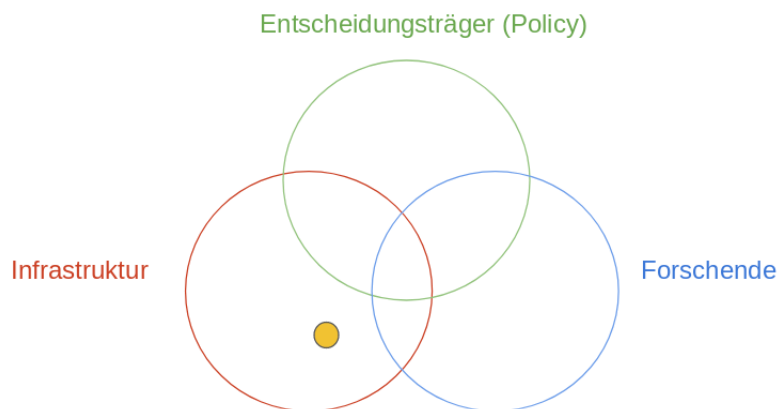


Abbildung 53: Workflow set-up and provenance information management. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	

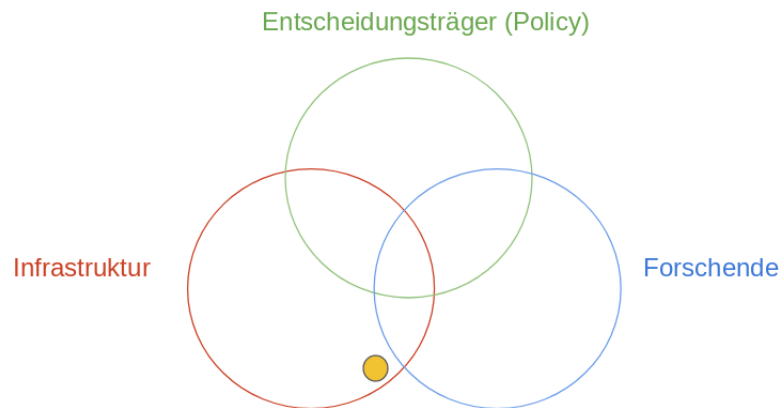


Abbildung 54: Cloud environment and storage management. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	

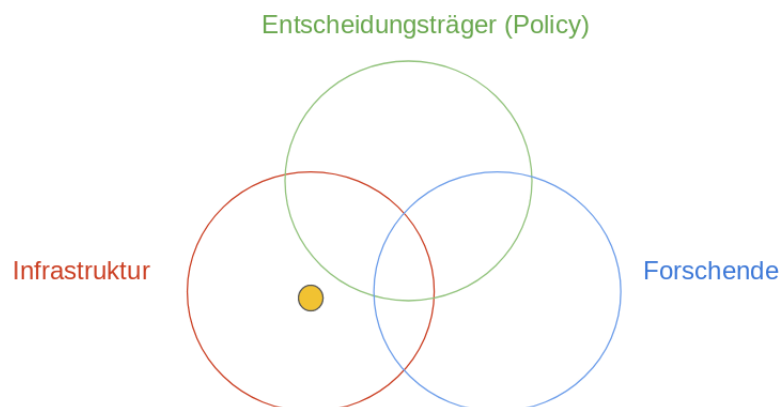
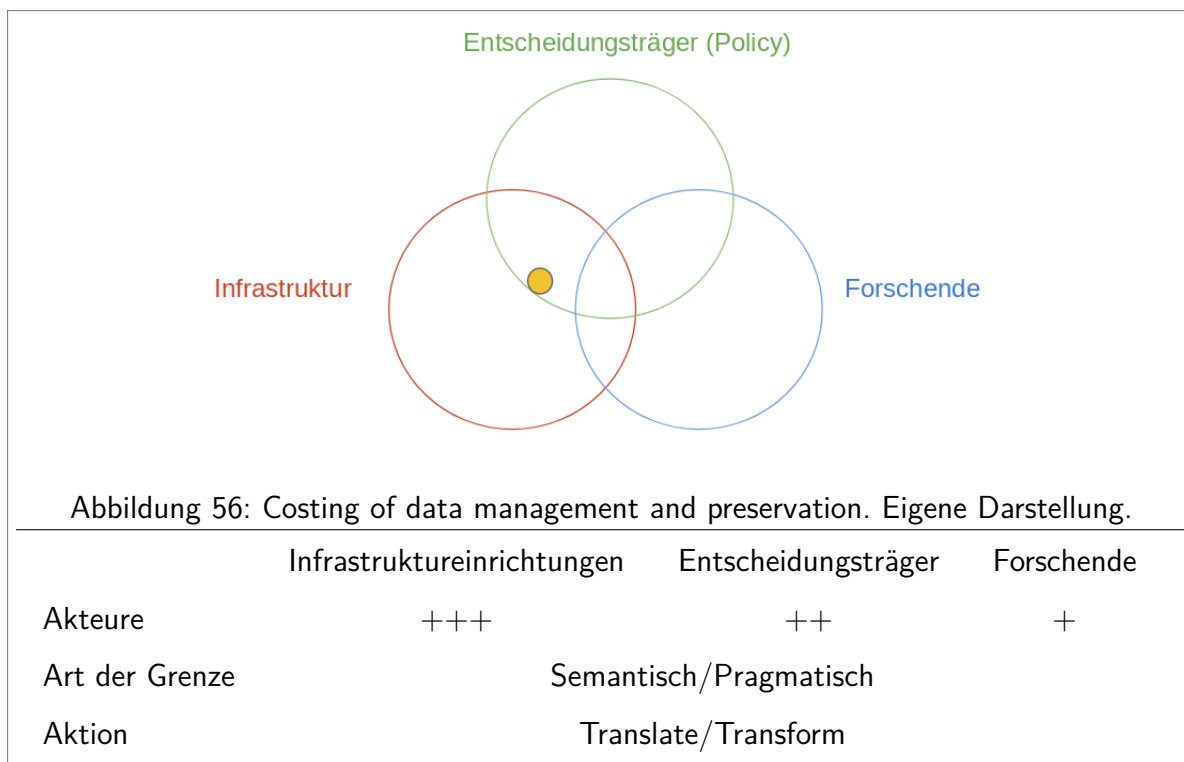
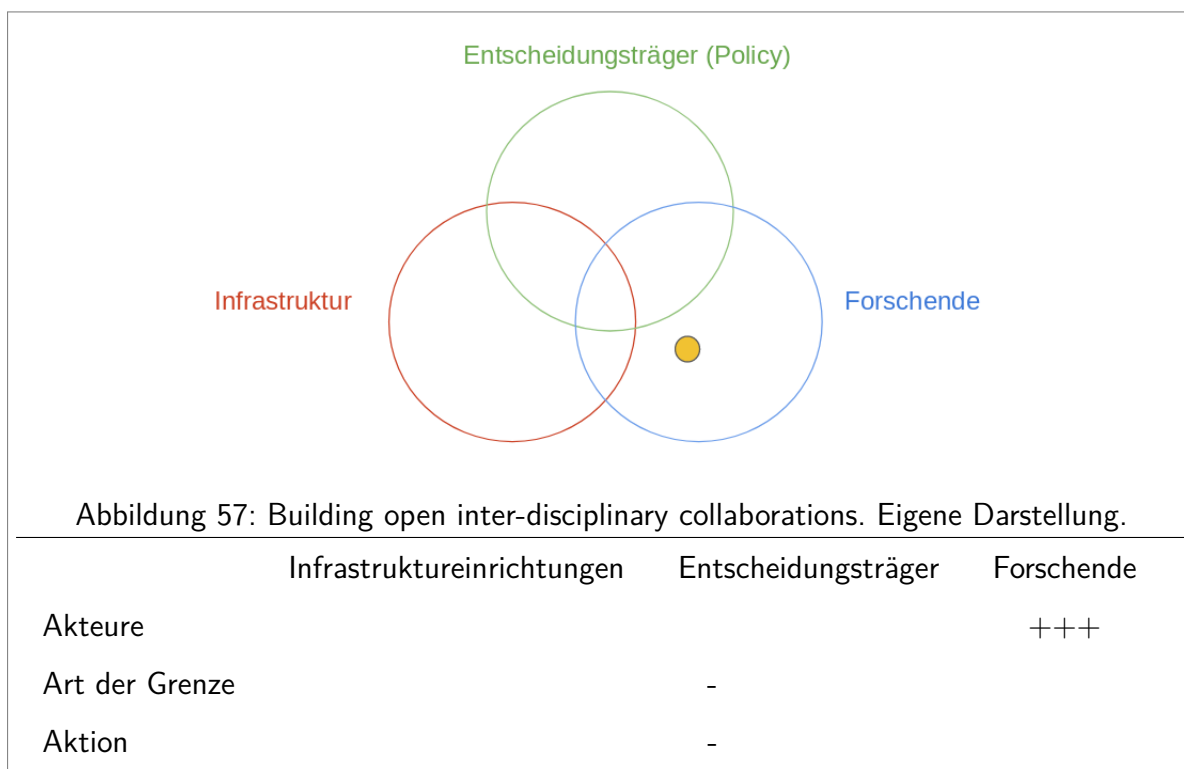


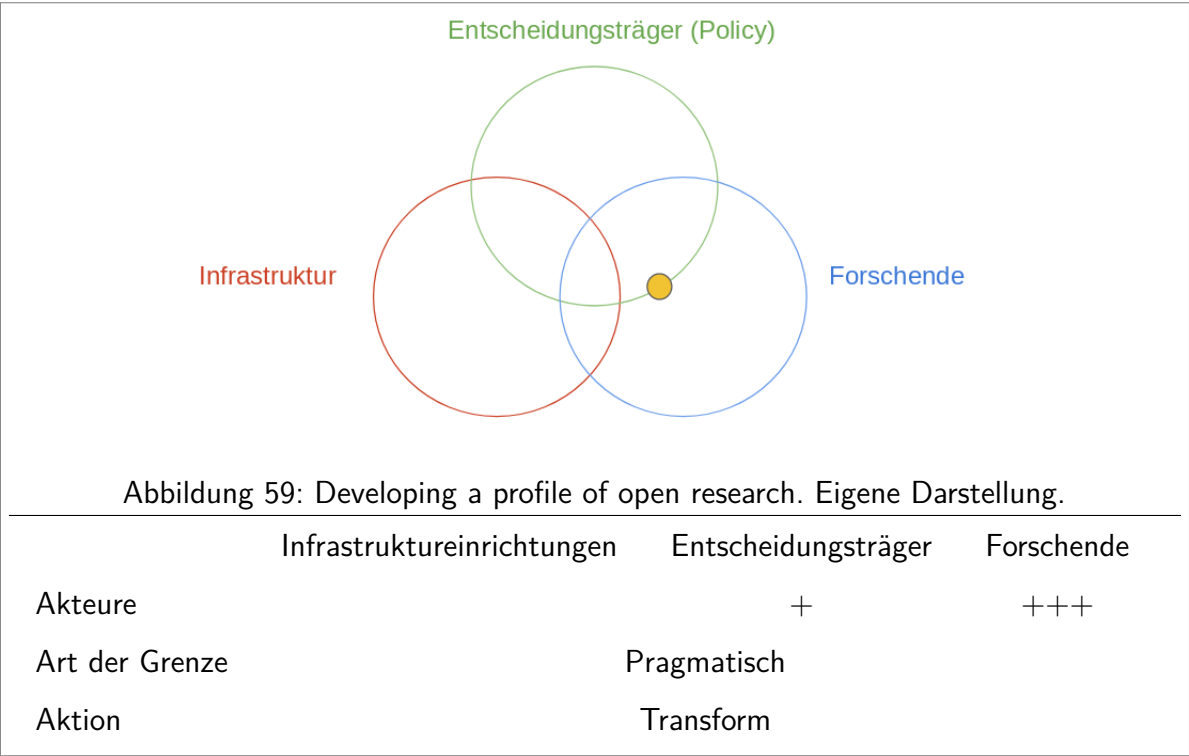
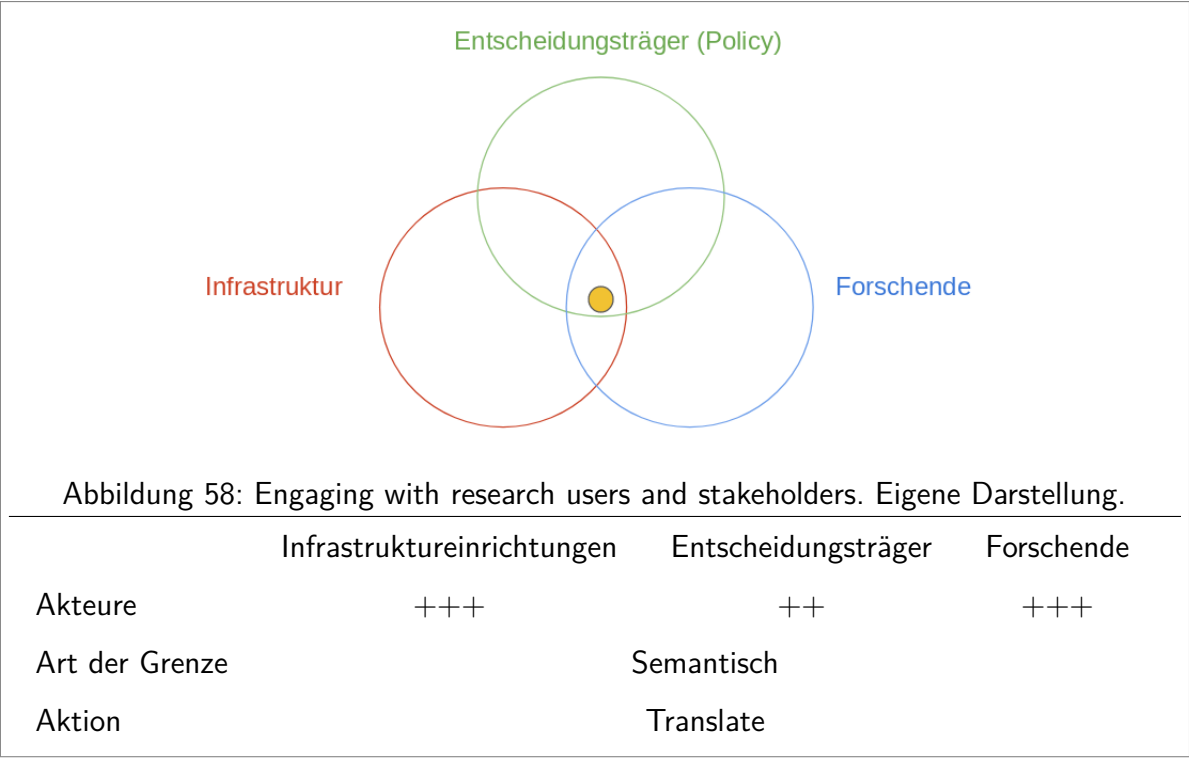
Abbildung 55: Authentication and authorisation (AAI) management. Eigene Darstellung.

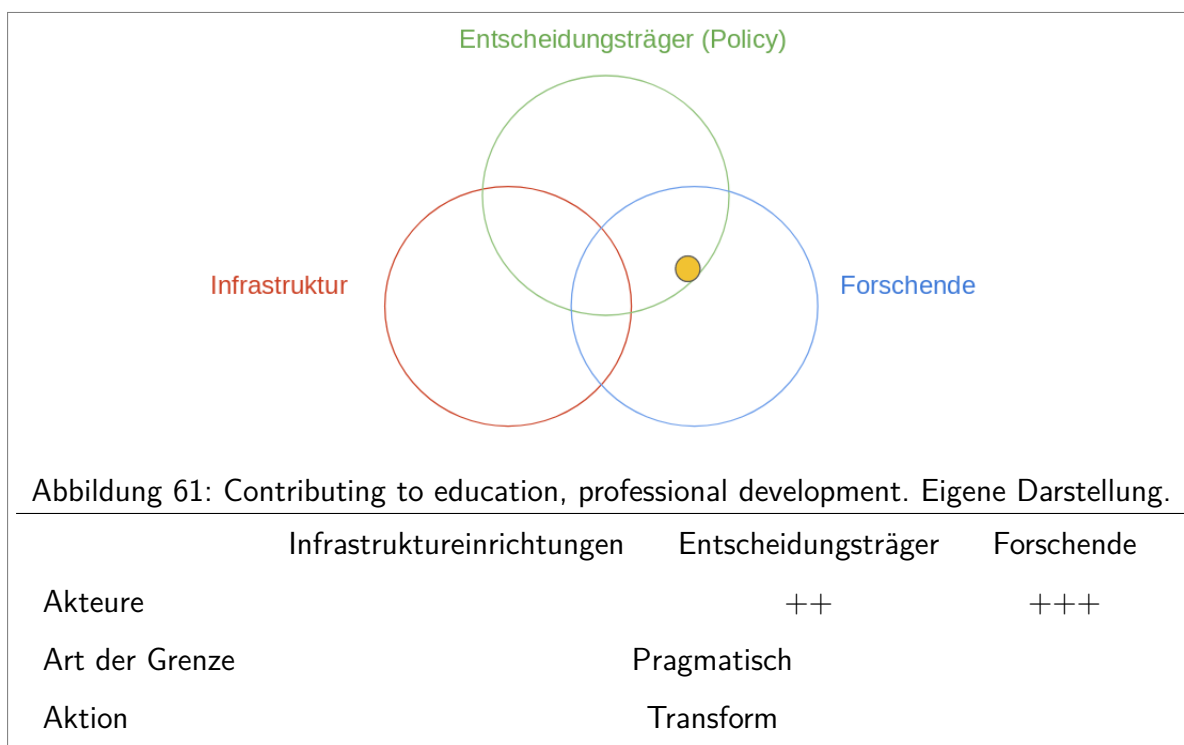
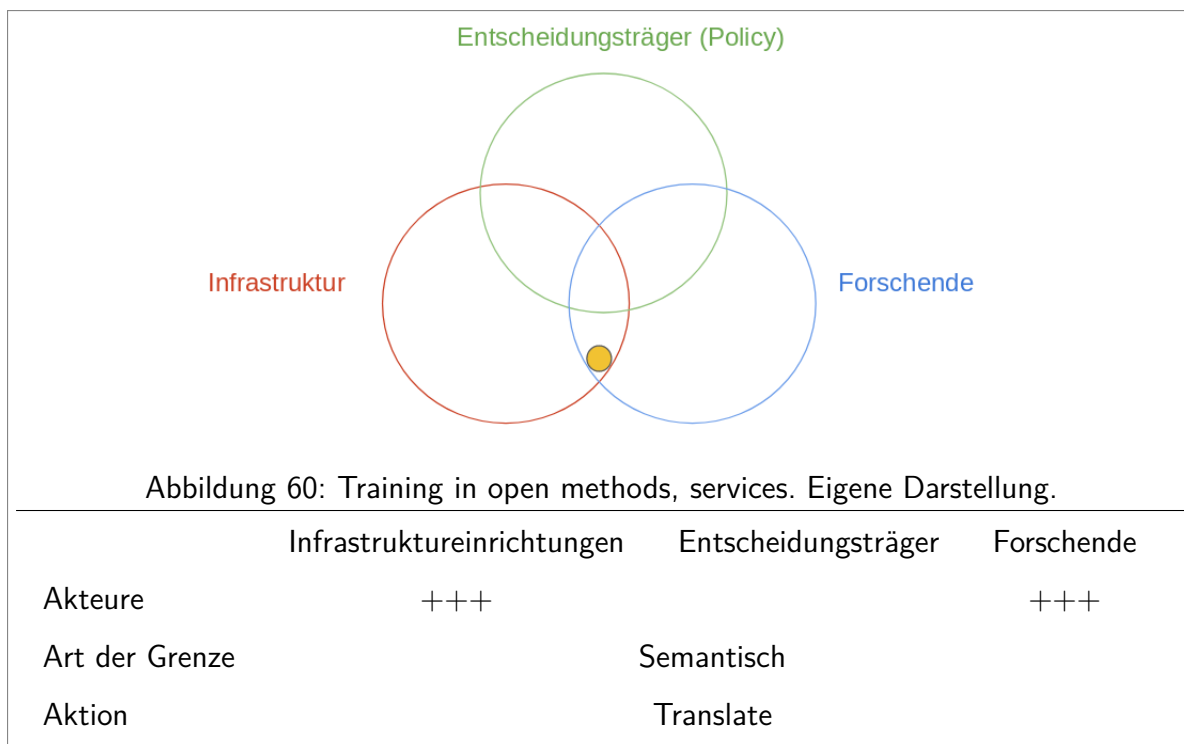
	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	



A.9 Advise and enable







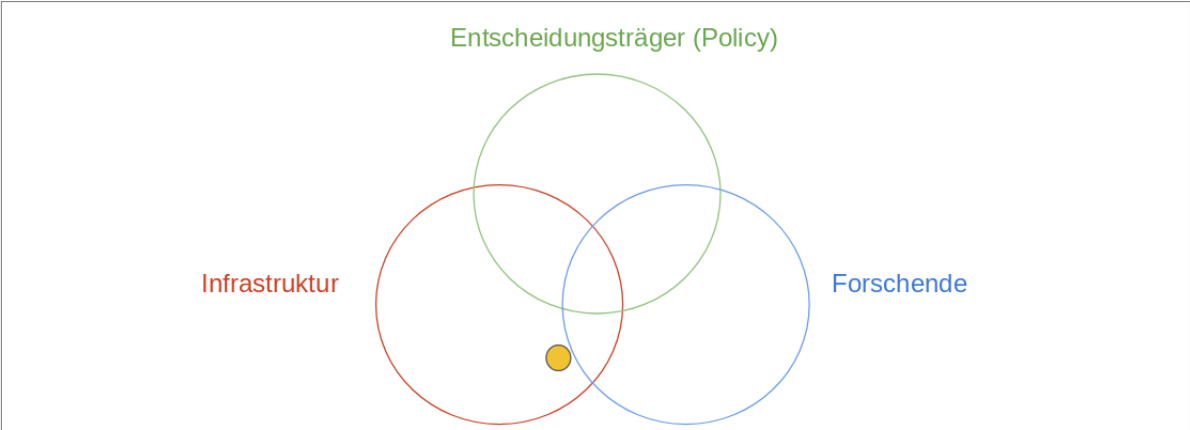


Abbildung 62: Contributing to open RES, networks, standards bodies. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure	+++		
Art der Grenze		-	
Aktion		-	

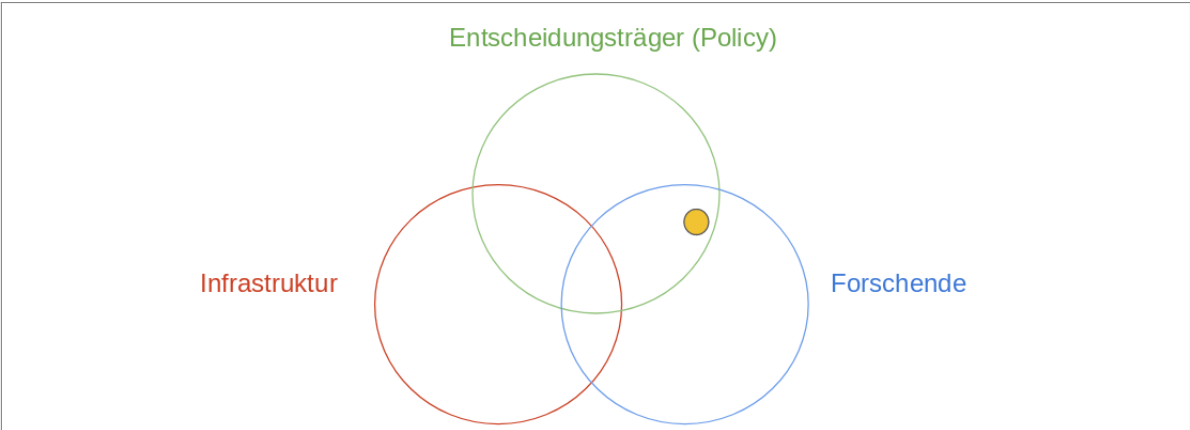


Abbildung 63: Supervision and mentoring. Eigene Darstellung.

	Infrastruktureinrichtungen	Entscheidungsträger	Forschende
Akteure			+++
Art der Grenze		-	
Aktion		-	